MECHANISTIC INSIGHTS INTO EUKARYOTIC TRANSCRIPTION USING NOVEL CRYSTALLOGRAPHIC TECHNIQUES

by

Christopher O'Neil Barnes

Bachelor of Science in Chemistry, University of North Carolina, 2008

Bachelor of Arts in Psychology, University of North Carolina, 2008

Masters of Arts in Chemistry, University of North Carolina, 2010

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Christopher O'Neil Barnes

It was defended on

December 2, 2016

and approved by

Angela Gronenborn, PhD, Professor of Structural Biology

Bruce Freeman, PhD, Professor of Pharmacology and Chemical Biology

Karen Arndt, PhD, Professor of Biological Sciences

William Furey, PhD, Professor of Pharmacology and Chemical Biology

Dissertation Advisor: Guillermo Calero, PhD, MD, Assistant Professor of Structural Biology

MECHANISTIC INSIGHTS INTO EUKARYOTIC TRANSCRIPTION USING NOVEL CRYSTALLOGRAPHIC TECHNIQUES

Christopher O'Neil Barnes, PhD

University of Pittsburgh, 2016

DNA-directed RNA Polymerase II (Pol II) is highly conserved among eukaryotic organisms and plays a fundamental role in cellular life, specifically gene transcription. This process is universal and is at the core of gene regulation; therefore, understanding its molecular details will provide essential clues that could potentially lead to pharmacological manipulation of gene expression.

In this work we reveal, for the first time, the nature of interactions between Pol II and a complete nucleic acid scaffold. The upstream double helix lies over a wedge-shaped loop from $Rpb2^{862-874}$ that engages the minor groove, providing part of the structural framework for DNA tracking during elongation. Rudder and fork loop-1 residues insert in between the DNA/RNA hybrid and non-template strand at the upstream end of the bubble, suggesting a direct role in coordinating annealing of DNA strands. At the downstream end, interactions between the non-template strand, $Rpb1^{1103-1112}$ and $Rpb2^{501-510}$ support an open fork. These Rpb1 residues form a rigid domain with the trigger loop (TL), stabilizing its off, non-catalytic state, indicating that "on/off" conformational transitions, where the TL moves in close proximity to the matched nucleotide for catalysis, may be linked to interactions with the non-template strand.

To further examine the role, structurally, of the trigger loop and bridge helix during translocation and catalysis, we characterized the gain of function T834P variant of Pol II. Structural studies of T834P in the presence of a matched nucleotide revealed the complete TL in

the on-state, a conformation only observed previously in two low resolution datasets. Our preliminary analysis suggests that kinking of the bridge helix potentially shifts equilibrium into the post-translocation state during the nucleotide addition cycle, which would prolong interactions of the trigger loop with the matched nucleotide in the insertion site. Elucidation of these structures was achieved through the use of X-ray free electron laser sources, which extended the resolution to 3.3 Å. Methodologies, like serial femtosecond crystallography and transmission electron microscopy, to aid in structural studies of highly relevant, and challenging protein targets, including multi-protein complexes of the transcriptional machinery are also discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATION

This work is dedicated to my two sons,

Julius and Maximus,

for the joy and love you provide on a daily basis.

ACKNOWLEDGEMENTS

During my graduate work, the Lord has blessed me with the opportunity to work and interact with highly skilled faculty, colleagues, and other students, while enjoying the support of many others along the way.

I would like to offer my sincerest gratitude to my mentor Dr. Guillermo Calero. His general enthusiasm for research and contagious energy is truly inspiring. He is an exemplary researcher, whose skills in structural biology are remarkable. I cannot count the number of times that I sat in his office and watched him refine low resolution structures in total amazement at his ability to visualize the protein backbone. Dr. Calero taught me a plethora of techniques and the importance of collaboration, all while exposing me to up and coming technologies that will reshape the structural biology field for years to come. He helped position me for a future career in science, yet that pales in comparison to his friendship and support throughout my time in his lab when I needed it most. For always being there and reminding me to be creative, I am forever grateful and hope that we will maintain this relationship into the future.

I would also like to thank the talented group of people that I have had the honor of working and learning from. Without their help, most of this work would not have been possible. In particular, I would like to thank Dr. Guowu Lin and Dr. Hilary Stevenson for their assistance in many experiments, and their continued friendship the last five years. I would also like to thank Dr. Filippo Pullara for teaching me the inner workings of the Calero lab, and for helping me to

solve numerous problems during my initial attempts at Pol II and TFIIF purifications. I would also like to thank the rest of the Calero Lab, past and present, who helped me along the way and made my life a little bit easier through their support and friendship.

During this work, we have also received help from many collaborators, who have contributed so much to the advancement of our projects. I would specifically like to thank Dr. Aina Cohen and Dr. Craig Ogata for their assistance in data collection and training during my trips to Stanford and Chicago to collect X-ray diffraction data. In addition, to my colleague Dr. Marcin Paduch, who offered assistance and expertise in the development of over 70 synthetic Fabs, and Dr. Craig Kaplan who supplied biochemical support for all yeast Pol II experiments.

In addition, I would like to thank my committee members, Drs. Angela Gronenborn, Bruce Freeman, Karen Arndt, and Bill Furey for investing their time and effort into meaningful discussions on how to stay focused and improve my work. Moreover, I'd like to thank the individuals involved with Molecular Pharmacology program, specifically the directors, Drs. Patrick Pagano and Guillermo Romero, for their continued support and belief in my abilities throughout the years. To all the administrative and faculty whom I did not name, I thank you for encouraging and supporting me and my family, particularly the staff at University Child Development Center who kept our kids safe, happy and fed while I worked towards my degree.

Last but definitely not least, I would like to thank my family, friends and Naima for their love, support and care. Raising two small children with the rigors of graduate school is definitely a challenge, but one that was made much easier alongside Dr. Naima Sharaf. I am beyond blessed to have you in my life, and the love of our families to help us along the way. Thank you to our moms, Vanessa Barnes and Rocio Palma, for the unreturnable gift of time, to come and take care of us and watch after Julius and Max when we had no daycare. Thank you to my dad,

Will Barnes, brother Alex, and sister Dominique for encouragement and reminding me to stay focused. You have all taught me the importance of hard work, dedication, and perseverance in the face of adversity. I cannot thank you enough for all that you mean to me and I hope someday I can return, with interest, the love you have shown me.

ABBREVIATIONS

RNAP	RNA Polymerase
Pol II	RNA Polymerase II
S. cerevisae	*Saccharomyces cerevisiae*
NAS	Nucleic Acid Scaffold
NAC	Nucleotide Addition Cycle
TF	Transcription factor
GTF	General Transcription Factor
NTS	Non-Template Strand
TS	Template Strand
TIC	Transcribing Intermediate Complex
TC	Transcribing Complex
TB	Transcription Bubble
Δ4,7	RNA Polymerase II Δ4Rpb4/7 subunits
NTP	Nucleotide Triphosphate
ATP	Adenosine Triphosphate
GTP	Guanosine Triphosphate
FL1	Fork Loop 1
FL2	Fork Loop 2

BH	Brige Helix
TL	Trigger Loop
U-loop	Rpb1 residues 1103-1109
TLB	Trigger Loop Bundle
TEM	Transmission Electron Microscopy
MPC	Multi-protein complex
NC	Nanocrystal
microED	Micro-electron diffraction
cryoEM	cryo electron microscopy
SFX	Serial Femtosecond Crystallography
XFEL	X-ray Free Electron Laser
SSRL	Stanford Synchrotron Radiation Lightsource
SLAC	Stanford Linear Accelerator Center
LCLS	Linac Coherent Light Source
CXI	Coherent X-ray Imaging
CSPAD	Cornell-SLAC hybrid Pixel Array Detector
XPP	X-ray Pump Probe
MX	Macromolecular Crystallography
E. coli / Ec	*Escherichia coli*
dGTPase	Deoxyguanosine triphosphate phosphohydrolase
HPCD	Homoprotocatechuate 2,3-dioxygenase
GOF	Gain of function
LOF	Loss of function

1.0 INTRODUCTION – EUKARYOTIC TRANSCRIPTION AND THE BIG PICTURE

Discovery of DNA's double-helical structure in the mid-1950s led to the postulation, by Francis Crick, of a central dogma that governs molecular biology (*1*, *2*). This doctrine explains the transformation of the coded hereditary information inside DNA into RNA molecules, the latter of which encodes a readable program for the synthesis of proteins by the ribosome (*3*). While exceptions to this classical view have been touted in recent years through the discovery of functional RNAs and viral enzymes capable of reverting RNA back to DNA (*4*-*7*), the decoding of DNA to RNA is at the root of cellular function and survival. This universal process, termed transcription, is driven by RNA polymerases (RNAP) which are able to generate RNA from a DNA template, independent of a primer.

While prokaryotes are capable of synthesizing all cellular RNA by a single polymerase, eukaryotic genomes code for three core polymerases: RNA Polymerase I (Pol I), II (Pol II), and III (Pol III). RNA polymerase I is responsible for the production of ribosomal RNA (rRNA), specifically the 5.8S, 18S, and 28S rRNA precursors, while 5S rRNA, transfer RNAs (tRNA) and small non-coding RNAs, are transcribed by Pol III. In addition to transcribing many small non-coding RNAs, the transcription of ~25,000 protein-coding genes in the human genome into messenger RNA (mRNA) is achieved by Pol II. Despite differing in function, promoter specificity, and localization, the 10-subunit core architecture of these enzymes is highly

conserved (*8*). Furthermore, conservation extends to the structure and functionality of a myriad of proteins, including the general transcription factors (GTFs), necessary for the recruitment of polymerases at DNA promoter elements.

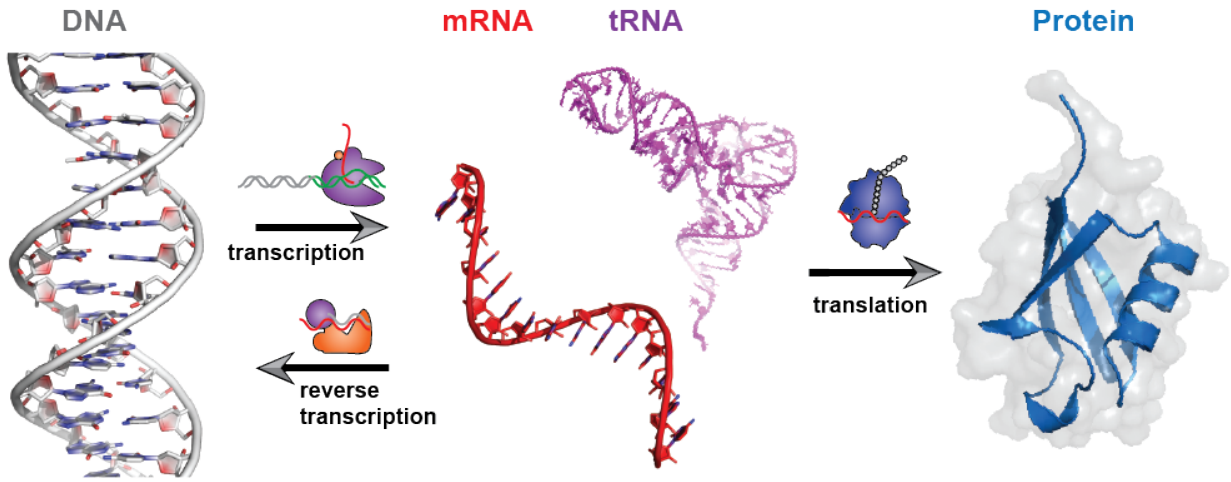

Figure 1. The central dogma of molecular biology

The flow of genetic information from DNA to RNA to Protein (classic view) is essential for cell function & survival.

In a coordinated effort, RNA polymerases engage transcription factors (TFs) and a host of nuclear proteins to either promote or inhibit transcription. In the context of Pol II, unregulated transcription can lead to aberrant gene expression patterns which potentially manifest in disease phenotypes. Indeed, there exist a copious number of disease-driven proteins that involve TFs or indirectly modulate TF activity (*9*, *10*). For instance, recruitment of Pol II to the promoter by the general transcription factor IIB (TFIIB) and interactions with TFIID (including the TATA-binding protein) are initial steps in transcription and represent the first chance at regulating gene expression. Biochemical studies have implicated TFIIB as a binding partner for many proteins associated with disease states. Specifically, both viral (Hbx, Vpr, EB1) (*11-13*) and oncogenic (Sp1, NFκB) (*14*, *15*) trans-activators bind TFIIB enhancing its ability to recruit Pol II to begin transcription. Moreover, overexpression of TFIIB or the B-related factor Brf2 (utilized by Pol

III) is significantly correlated with tumor angiogenesis and poor patient survival in human hepatocellular and esophageal squamous cell cancers, respectively (*14-18*). Furthermore, the implication of Mediator, a multi-subunit complex that bridges TF and Pol II activity, as an instigator of cancer and developmental disease (*19*) begs the question: can we target transcription machinery in the treatment of human disease?

This idea has remained an open question due to the lack of natural ligands and the challenges of targeting specific protein-protein or protein-nucleic acid interactions with traditional small molecules. Yet, in recent years, the use of structural information to rationally design therapeutic agents to target transcription machinery (ie. the p53-MDM2 complex or c-Myc activity) has made the once undruggable, druggable (*10, 20*). However, to continue this trend in exploiting TFs as therapeutic targets, an increase in the detailed structural and biochemical knowledge of the interactions that govern their activity and stability is required. In this work, we provide detailed structural insights into the binding mechanism of yeast RNA Polymerase II with a complete nucleic acid scaffold in the presence of GTFs. Additionally, in later chapters we explore emerging techniques in structural biology to aid in the collection of high-resolution data of multi-protein complexes, which we applied to multiple Pol II – TF complexes in the presence of nucleic acid scaffolds. Increasing our understanding of the molecular details that govern transcription will provide essential clues that potentially lead to pharmacological manipulation of gene expression in disease states.

2.0 STRUCTURAL BASIS OF EUKARYOTIC TRANSCRIPTION

Traditionally, x-ray crystallography has been the method of choice for protein structure determination, especially for large molecular weight proteins (*21*). Starting in the early 1990s, the first detailed structures of RNA Polymerases were revealed by examining the free enzyme of the single-subunit bacteriophage T7 RNAP (*22*), followed by the characterization of an elongating RNAP a few years later (*23*). While unrelated in both sequence and structure to multi-subunit polymerases, the catalytic mechanism which governs nucleotide addition provided the first insights into transcription. It wasn't until the next decade that the first atomic details of multi-subunit polymerases became available when the structures of the yeast and bacterial RNA polymerases were revealed (*24*, *25*). In this chapter, our understanding of the mechanistic and structural details of *Saccharomyces cerevisiae* Pol II transcription will be discussed along with a brief examination of the techniques used to collect the experimental data.

2.1.1 Structure of *S. cerevisiae* Pol II: a brief history

In contrast to the single-subunit T7 RNAP, structural studies of eukaryotic polymerases are a challenging endeavor because of their size and complexity. As shown in Figure 2, since their initial discovery in the late 1960s, almost three decades would pass before the first structural information could be collected. An initial hurdle to the study of eukaryotic polymerases was the

separation of an active enzyme from eukaryotic cells with the capacity to transcribe purified DNA (*26*, *27*). Fortuitously, the isolation of a transcriptionally active form of Pol II from yeast in the 1980s, by the lab of Roger Kornberg, paved the way for extensive biochemical and structure-function studies of the eukaryotic transcription machinery (*28*). Over the next decade, a series of attempts were made to crystallize Pol II. From these early two-dimensional crystallization experiments using lipid scaffolds two important conclusions were drawn. First, sample heterogeneity, due to sub-stoichiometric amounts of the heterodimer Rpb4/Rpb7, had deleterious effects on crystal order and growth (*29*, *30*). Second, the use of a deletion strain (*31*, *32*) which eliminated the Rpb4/7 heterodimer (PolΔ4/7) not only improved crystal growth, but also provided evidence that Pol II transitions between two distinct forms, an open/closed state necessary for the entry and retention of DNA, respectively (*33*, *34*).

From these early two-dimensional crystallographic results and a revamped large-scale fermentation protocol to natively purify the quantities of protein necessary for *in vitro* experiments (*35*), Fu *et al*. were able to promote Pol II crystallization in traditional vapor diffusion trials for single crystal x-ray diffraction experiments (*36*). Their findings, using a 6Å phased map derived from heavy atom clusters, led to a more robust understanding of cleft and clamp domain movements, but the atomic details of Pol II remained elusive. A breakthrough finally occurred in the following years when it was discovered that a special *in situ* dehydration treatment of Pol II crystals resulted in the shrinkage of the asymmetric unit that ultimately decreased non-isomorphism and improved crystal contacts along the clamp domain (*37*). As a result, atomic details of the 10-subunit Pol II core in the absence (*25*) and presence of a DNA/RNA hybrid (*38*) were revealed to resolutions of 2.8Å and 3.3Å respectively, providing for the first time structural insights into eukaryotic molecular machines, and paving the way for

future studies of transcription (Figure 2).

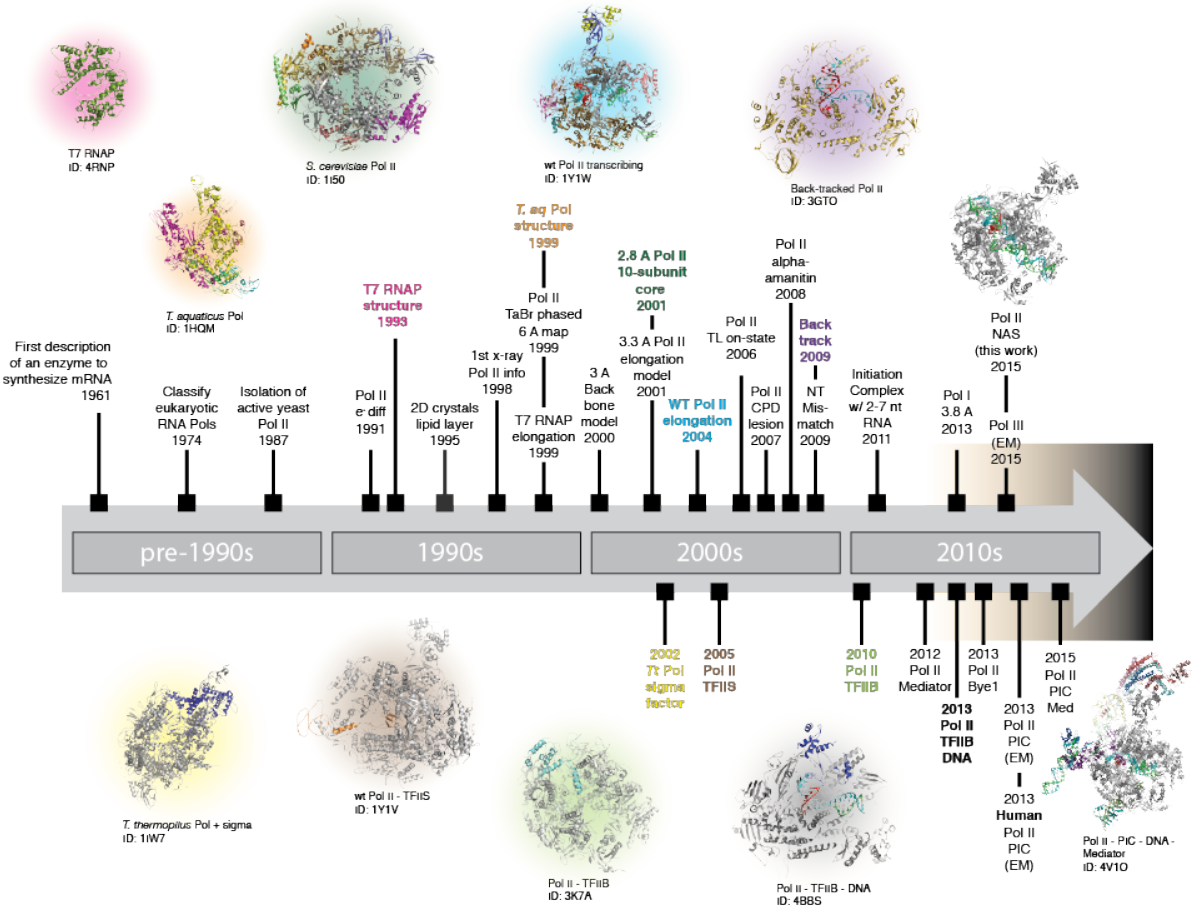

Figure 2. Timeline of structural insights into transcription

Timeline highlighting high-resolution structures which added to our understanding of transcription. Cartoon representation of structures were rendered in PyMOL using the referenced PDB ID. Models are color-coded to match their point in history, while the box at the end of the timeline represents the current period driven by the "resolution revolution" in single-particle cryo-EM (section ***2.1.5).***

2.1.2 Free and elongating structures of Pol II

At the time, these structures were an amazing feat considering Pol II's molecular weight of >500 kDa, along with the numerous technical and phasing challenges that were overcome to arrive at an accurate model (*39*, *40*). What these combined structures revealed, in addition to the

prokaryotic Pol structure solved from *Thermus aquaticus* (*24*), was a common architecture by which the two largest subunits, Rpb1 and Rpb2 (comparable to ß and ß´ in bacterial Pol), formed a positively charged cleft capable of binding DNA. Along the periphery, smaller subunits contribute to core assembly and stabilization, with subunits Rpb3 and Rpb11 having similar homologs in bacterial Pol. Interestingly, Cramer *et al*. were able to capture Pol II in two different open states, driven by mobility in the clamp domain of Rpb1, compared to the closed state induced by the presence of a tailed DNA/RNA hybrid in the active center (*38*). This transition from open to closed states is mitigated by five "switch" regions which show conformational plascity in the absence of nucleic acids, but couple clamp closure upon DNA/RNA binding that helps to stabilize elongating complexes.

In the years directly following these structures, two groups successfully captured the complete 12-subunit Pol II (*41*, *42*). Interpretations of the models presented by both authors were virtually identical, as it was found that the Rpb4/7 sub-complex plays a role in Rpb1 clamp dynamics. Protruding from outside the core, the Rpb7 "tip" domain wedges between the clamp and the linker of Rpb1 C-terminal domain (CTD) inducing a conformational change in two switch regions that promote a closed Pol II state. For this reason, it was suggested that DNA loading occurs after melting above the cleft by other TFs since the closed clamp state prevents duplexed DNA from entering the cleft.

Building on the results of Gnatt *et al*. which showed that the incoming DNA duplex bends some 90˚ upon reaching the active center, several transcribing structures were solved using synthetic tailed DNA/RNA hybrids (*43*, *44*), as well as mimicked transcription bubbles (*45*) to provide insight into the nature of the interactions between Pol II and nucleic acids as shown in Figure 3.

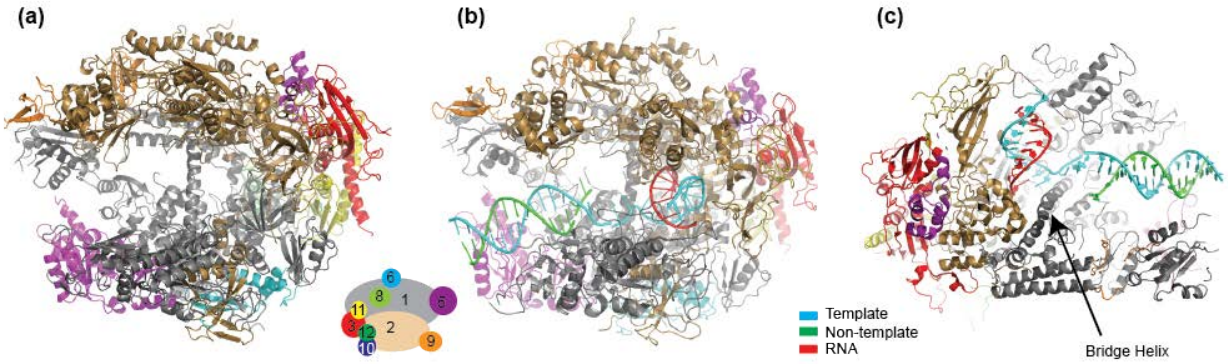

Figure 3. Structure of free and elongating 10-subunit Pol II

(a) *Cartoon representation of 10-subunit Pol II (PDB:ID 1I50), with subunits color coded based on schematic below.* ***(b,c)*** *Top and side view, respectively, of elongating Pol II (PDB:ID 1R9T) with nucleic scaffold color coded as template strand (cyan); non-template strand (green); RNA transcript (red).*

Collectively these structures show that the DNA is cradled between the two largest subunits through a larger network of H-bond driven interactions. The cleft is able to accommodate a 12-18 single-strand DNA template in the transcription bubble (TB), which is then free to base pair with the incoming nucleotides to form the basis of the DNA/RNA hybrid at the back of the cleft (*38*). The 3´- end of the nascent RNA transcript stays in contact with the active center, particularly the bridge helix domain that runs perpendicular to the DNA duplex. The 5´end of the RNA separates from DNA template around position i-7 to i-9 (where the i+1 is the addition site of the incoming nucleotide, and i-1 site is the position of the most recently added NTP), and is directed towards the exit tunnel (*44*, *45*).

2.1.3 Mechanism of nucleotide selection, catalysis, and translocation

During elongation, Pol II catalyzes the reaction of an incoming nucleotide to the growing mRNA transcript in a multi-step process termed the nucleotide addition cycle (NAC), the basis of which

was identified first in T7 RNAP (*46*), followed by studies in bacterial (*47*) and eukaryotic polymerases in the presence of matched nucleotides in the A site (*44*, *45*, *48*, *49*). To start, a post-translocated elongation complex provides access to the i+1 site for nucleotide base-pairing with the DNA template. The matched nucleotide is trapped by two highly conserved domains of RNA polymerases, the bridge helix and trigger loop (which will be discussed in more detail below). Discrimination of the correct nucleotide by residues on both domains and Pol II proximal domains induces folding of the trigger loop, which provides the catalytic residues for nucleotide addition.

The addition of the nucleotide uses a two-metal ion mechanism as previously proposed from T7 RNAP experiments (*46*) and visualized in Figure 4A. Metal A in the active site is permanently bound, coordinated by three aspartic acid residues of Rpb1, whereas the Metal B is mobile and can bind the incoming nucleotide at either the β/γ phosphate (*49*). Pol II catalyzes the addition of the incoming nucleotide via a nucleophilic substitution mechanism (S_N2), with the 3´-hydroxyl group on the nascent RNA acting as the nucleophile that attacks the α–phosphate of the NTP. After addition of the nucleotide to the 3´end of the RNA chain, release of a pyrophosphate ion (PP_i) occurs before translocation across the nucleic acid scaffold to reach the next addition site.

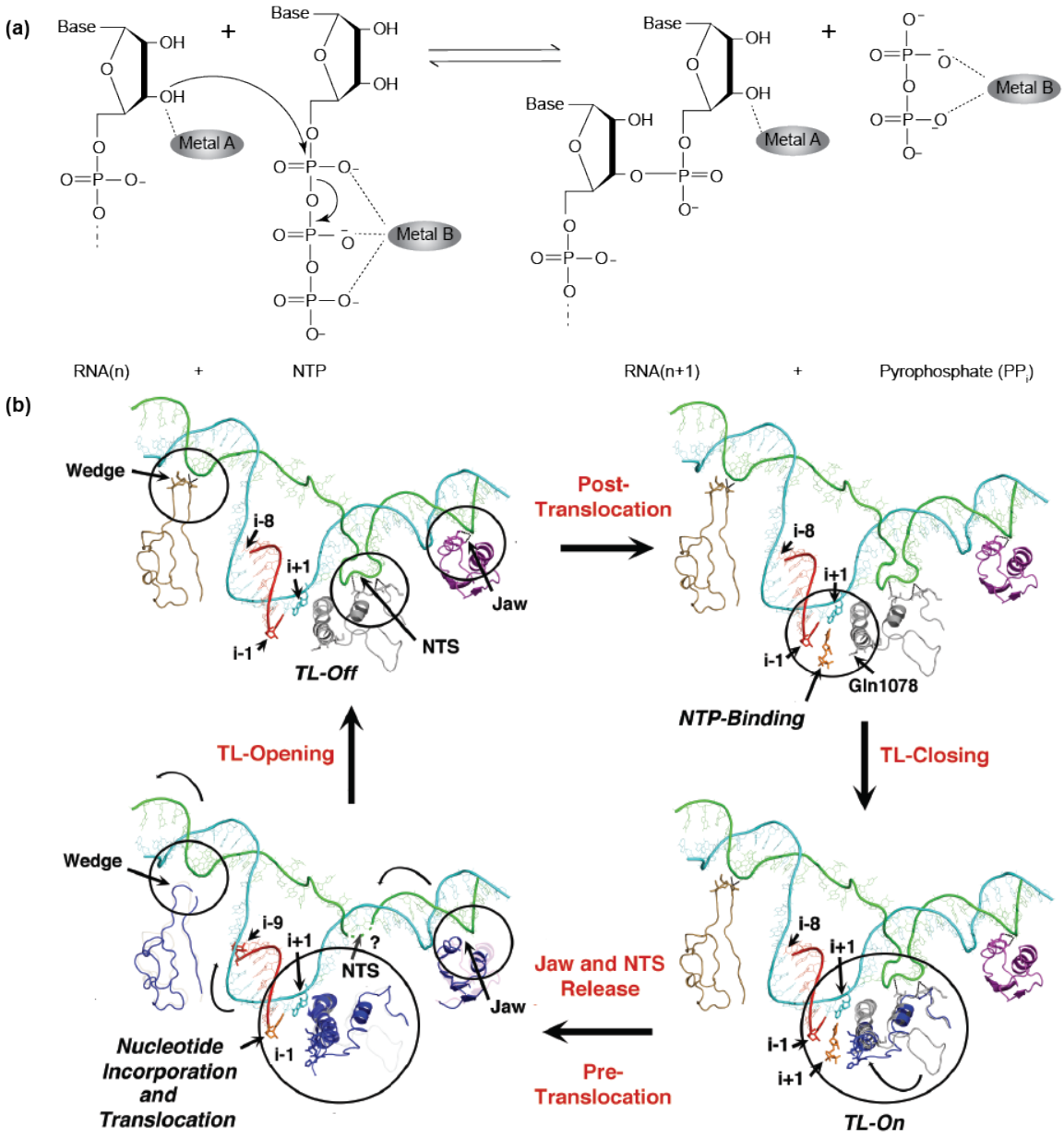

Figure 4. Nucleotide catalysis and TL driven translocation

(a) *Reaction mechanism of nucleotide catalysis in RNA polymerases.* ***(b)*** *A working model of nucleic acid scaffold translocation. Left upper panel: After nucleotide addition, TL retreats to off-state conformation. Right upper panel: A matched nucleotide binding to the post-translocated state (TL "off" conformation) is recognized by Gln1078. Right lower panel: high affinity interactions of a matched NTP result in morphing of the TL from "off" to "on" states. Left lower panel: NTP is incorporated into the nascent transcript (nucleotide addition), release of NAS interactions with jaw domain and proximal Pol II domains facilitate its translocation.*

The resulting complex is now in a pre-translocated state, with a matched NTP residing in the insertion site. Thus, Pol II must translocate to the i+2 site to reveal the insertion site to repeat the NAC. To achieve this a number of interactions between Pol II and the nucleic acid scaffold (NAS) must be broken, which will be discussed in greater detail in Chapter 3. As mentioned above, however, the BH and TL play a critical role in NAS translocation. The BH is a structurally conserved domain comprising residues 815-845 of Rpb1 in yeast Pol II. Biochemical and structural data in bacterial polymerases (*50*, *51*) suggested that the conformational "bending" and straightening of the BH underlies translocation. Indeed, in archaeal systems the bending of the bridge helix is critical for catalysis, since mutations to introduce kinks led to super active polymerases (*52*) which will be explored further in this work.

Further biochemical and mutational studies showed that BH and TL residues function cooperatively during the NAC and subsequent translocation (*53*-*58*), leading to a proposed mechanism of a Brownian ratchet motion, in which an equilibrium between pre- and post-translocation states exist as rearrangements of TL and BH domains occurs (Figure 4B) (*53*). Detailed structural analysis revealed the nature of TL mediated translocation by inhibiting translocation with α–amanitin (*57*). This structure revealed that in the intermediate step of pre- to post-translocation, the TL occupies a "wedged" conformation, bending the central BH. This helped to validate the proposed mechanism of the Brownian ratchet, as α–amanitin inhibits Pol II by trapping the TL and BH movements, which impairs DNA translocation (*59*).

2.1.4 Understanding Pol II – TF interactions from high-resolution data: start site selection, backtracking and arrest

As technologies and biochemical methods improved, a determined effort began to understand the interplay between Pol II and the TFs that regulate its activity. The pre-initiation stage of transcription requires concerted interactions between RNA Polymerase II (Pol II) and the general transcription factors (GTFs) TFIIB, TFIID, TFIIF, TFIIE and TFIIH. Despite the plethora of TFs, few studies have examined the role of these factors in the context of a transcribing Pol II until very recently, which will be discussed in the next section. High-resolution structural studies using x-ray crystallography have been limited to a handful of Pol II complexes.

Structural studies of Pol II in complex with TFIIS (Figure 5A) were some of the first structural studies to show an additional factor bound. These studies revealed the nature of TFIIS's function in elongating complexes, particularly the ability of TFIIS to promote Pol II's weak nuclease ability during transcriptional arrest (*45*, *60-62*). This function is critical since Pol II has a tendency to oscillate between active and inactive states, driven by mismatched nucleotide incorporation (*62*, *63*), DNA lesions (*64-66*), or intrinsic properties of polyA regions of DNA, which exhibit weaker A:U base pairing. Typically, pausing events are short-lived and Pol II is capable of restarting elongation without the aid of additional factors (*67*, *68*), even when a few nucleotides are backtracked (*62*). However, prolonged pausing can lead to larger changes in the transcription register, resulting in arrest, which can only be overcome by TFIIS.

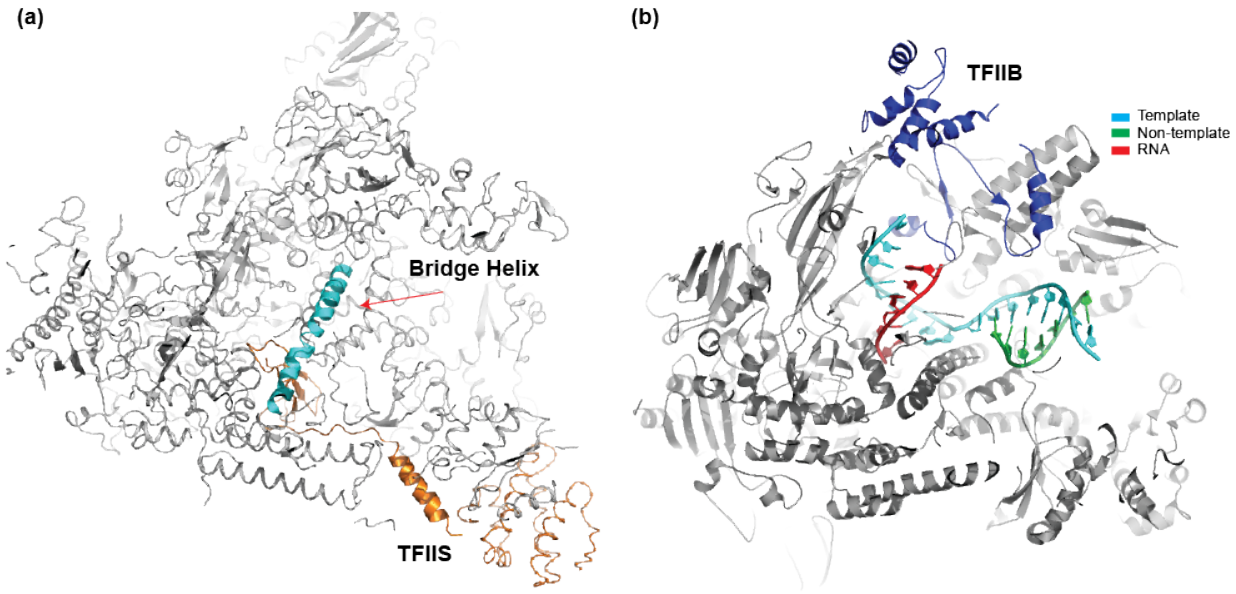

Figure 5. Architecture of Pol II - TFIIS and transcribing Pol II - TFIIB complexes

(a) *Crystal structure of Pol II – TFIIS (PDB:ID 1Y1V) showing how TFIIS (orange) threads into the active center, situating itself against the BH (cyan).* ***(b)*** *Structure of the transcribing Pol II – TFIIB (PDB:ID 4BBS) complex revealed interactions between TFIIB (blue) domains with Pol II and the nucleic acid scaffold.*

Crystal structures of Pol II backtracked structures revealed how the 3´end of the RNA enters the nucleotide funnel and active site pore, until the backtrack site is reached (*60*). TFIIS domains are capable of threading into the pore, thus locking the TL away from the active site where backtracked RNA resides. A conserved TFIIS hairpin domain, also observed in GreA/GreB in bacterial homologs (*69*), engages with charged residues near the site of backtracked RNA to initiate metal-dependent cleavage of RNA, using a water nucleophile. After RNA cleavage and release, a new 3´ end is formed, TFIIS is ejected, and elongation can resume.

Another factor that has been studied extensively, is the general transcription factor TFIIB. The role of TFIIB in transcription has been studied in greater detail than any other general transcription factor, as it has been shown to be indispensable for transcription initiation (*70-72*).

In eukaryotic systems, the core cyclin domains of TFIIB complexes with TATA-binding protein (TBP) on the promoter and interacts with upstream and downstream DNA (*73*). Once bound at the promoter, TFIIB acts as a bridge between promoter elements and Pol II utilizing its flexible zinc ribbon domain to recruit Pol II to the DNA providing directionality for transcription (*71*, *74*, *75*). Given its importance in the recruitment of Pol II, it is no surprise that TFIIB or TFIIB related factors are conserved among eukaryotic and prokaryotic polymerases (*8*, *76*, *77*). In addition, the B-finger/reader domains play a role in scanning for a conserved thymine in eukaryotic systems to determine transcription start site (TSS) (*78*).

To date, there exist four co-crystal structures of Pol II in complex with TFIIB, resulting in a detailed mechanism of how TFIIB functions in Pol II recruitment to DNA; transcription start site selection; and modulation of transcription within Pol II's active center (*74*, *79-81*). Briefly, the zinc ribbon domain (also termed the B-ribbon domain), engages Pol II's dock domain in Rpb1, while the cyclin domains interact with the wall of Rpb2. These extensive interactions with Pol II allow for the engagement of the B-finger and B-reader domains (highly conserved domains between the cyclin and ribbon domains) with the active site (Figure 5B) (*81*). The B-finger hairpin acts as a unidirectional extension of the B-ribbon and engages with the DNA/RNA hybrid at the base of the Pol II cleft. The B-linker helix locates against the NTS and may function in DNA opening or stabilization of the TB during early transcribing events. Moreover, the B-reader loop interacts with the TS to position the DNA for initiation by recognizing the TSS. In addition, contacts near the site of the RNA exit tunnel validates biochemical evidence that TFIIB assist in promoter escape after contacting the nascent 8-13 nucleotide RNA transcript and releasing the NTS for bubble collapse (*81*, *82*). Importantly, these models indicate that TFIIB induces active site rearrangements of the metal ions, by potentially increasing the affinity

for metal B and thus stimulating catalysis during the NAC (*81*).

Remarkably, of all the general transcription factors that assemble into the pre-initiation complex, only TFIIB and TFIIF dissociate from the promoter upon the transition from initiation to elongation and must be recruited again, along with Pol II, for subsequent transcription re-initiation (*78*, *83*). A complex involving Pol II and a small domain of TFIIF was only recently revealed using crystallographic methods (*84*), despite being tightly bound and recruited to the promoter in complex with Pol II. As a factor, TFIIF is required for accurate transcription initiation, since it may play a role in NTS stabilization in transcribing complexes (*85*). Yet, the minimal amount of high resolution structural information in context with Pol II limits our understanding of its function, along with the additional GTFs that have yet to be elucidated.

2.1.5 Characterization of the complete Pol II preinitiation complex

Despite breakthroughs in the reconstitution of the entire yeast preinitiation complexes (PIC) (*86*) coupled with chemical cross-linking techniques to identify flexible domains (*87*), efforts to obtain large, well diffracting crystals of Pol II in complex with multiple transcription factors failed. In recent years, the emergence of low-resolution single particle cryo-EM studies have significantly enhanced our understanding of the entire eukaryotic PIC (*84*, *85*, *88-93*) (Figure 6).

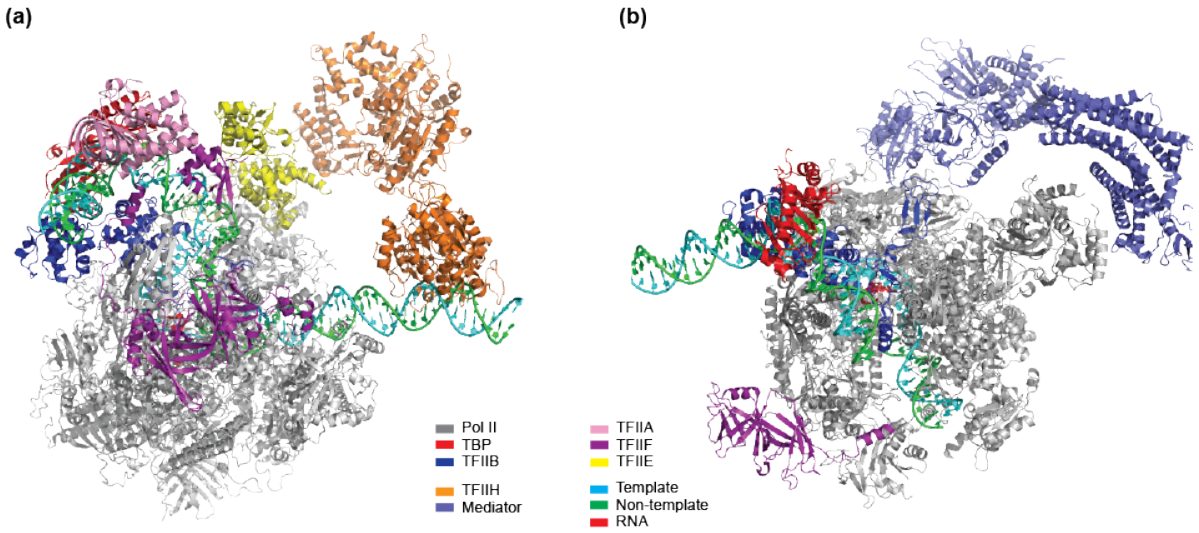

Figure 6. Cryo-EM models of Pol II pre-initiation complexes

(a) *The cryo-EM structure of the human PIC core (PDB:ID 5IY9), and* ***(b)*** *the yeast minimal PIC in the presence of Mediator (PDB:ID 4V1O). Docking of x-ray crystal structures allow for a pseudo-atomic modeling of the entire PIC.*

What these structures have revealed is the nature of PIC assembly at the promoter. Briefly, as described above Pol II can bind DNA non-specifically but cannot initiate transcription. Instead, Pol II requires a set of GTFs to initiate transcription in a process which occurs in three distinct stages: 1) Pre-initiation: TBP binds to a TATA-box (at the promoter site) 30 nucleotides upstream of the TSS and recruits TFIIB which in turns recruits Poll II-TFIIF and the TFIIH-TFIIE complex (Figure 6A) (*85*). Initial promoter melting is triggered by TFIIH helicases and generates a 7-9 nucleotide TB which passes through the Rpb1 rudder and Rpb2 fork loop 1 (FL1) residues towards Pol II's active center (*84*, *92*); 2) Early transcription: further unwinding to 18-25 nucleotides is possibly assisted by TFIIB or TFIIF followed by synthesis of the initial RNA transcript (*84, 90, 92*). 3) Elongation: The third stage marks the transition from initiation to elongation; transcripts of 10 or more nucleotides result in promoter escape and stabilization of a mature TB (TFIIF dependent) that ultimately leads to dislodging of the GTFs leaving a poised

Pol II for entry into productive elongation. In addition, recent structures have shed light onto the importance of Mediator, a multi-subunit complex, and its function in transcriptional regulation (*84*, *93*). As detector technologies continue to improve, higher resolution structures derived from cryo-EM may become attainable to help enhance our understanding of the overall architecture of the transcriptional machinery (*94*, *95*).

2.1.6 Open questions and the scope of this work

Structural research on Pol II has focused on 1) Pol II-DNA/RNA interactions, 2) Pol II in complex with TFIIB, and 3) most recently, the entire pre-initiation complex. Such structures have revealed the nature of the DNA-RNA hybrid, the mechanisms of nucleotide selection and catalysis, as well as the mechanism of backtracking. However, details of a complete nucleic acid scaffold (including a full transcription bubble) have remained elusive. Indeed, debate surrounding the number of nucleotides unwound in a mature bubble is still an important issue since sizes ranging from 8-22 nucleotides have been reported for bacterial, archaeal and eukaryotic polymerases (*82*, *96*, *97*). In addition, the size of the bubble itself might not be fixed but depend on Pol II transcriptional stage, as evidence of a scrunched state (where template and non-template strand bases are compacted in space relative to relaxed conformations) has been proposed for the early stages of transcription initiation in bacteria (*98, 99*). These questions were addressed in this work by co-crystallizing Pol II with a complete nucleic acid scaffold comprising various length transcription bubbles in the presence of TFIIF. Moreover, limited knowledge about DNA translocation in a transcribing Pol II was addressed by examining the network of interactions across the complete NAS with Pol II pre- and post- nucleotide addition.

Finally, despite the advances in single particle cryo-EM techniques, x-ray crystallography

still provides the best opportunity at garnering atomic-level details of the transcription machinery. Improvements in the ability to crystallize and collect data of large, multi-protein complexes (MPCs) is necessary. Therefore, one goal of this thesis was to optimize existing and novel crystallographic techniques for the study of MPCs, pertaining to the Pol II transcription machinery. To this end, transmission electron microscopy (TEM) techniques were developed to assay crystal quality and guide the optimization of Pol II crystallography experiments at conventional and novel x-ray free electron laser (XFEL) sources. Furthermore, unique multi-crystal holders were created for fixed-target XFEL experiments of Pol II complexes to improve data efficiency and quality.

3.0 CRYSTAL STRUCTURE OF A TRANSCRIBING RNA POLYMERASE COMPLEX REVEALS A COMPLETE TRANSCRIPTION BUBBLE

3.1 SUMMARY

The co-crystal structure of Pol II bound to a complete nucleic acid scaffold (NAS) reveals for the first time the architecture of the upstream double helix, the non-template strand (NTS) and the full-transcriptional bubble (TB). The upstream double helix lies over a wedge-shaped loop from Rpb2 (residues 862-874) that engages the minor groove of the duplex providing part of the structural framework for DNA-tracking during transcript elongation. Locations of the rudder and fork loop-1 (FL1), which form the arch that is inserted in between the hybrid and NTS, at the upstream end of the TB, suggest that residues involved in the contacts may play a direct role in coordinating annealing of the template and non-template strands. At the downstream (or opening) end of the TB, the interaction between NTS and a stretch of Rpb1 residues (1103-1112) apparently unwinds the downstream duplex DNA. These Rpb1 residues form a rigid domain with the Trigger Loop (TL, Rpb1 1078-1092), in which the TL is stabilized in the open state. These observations suggest that the conformational transition of TL (opening/closure) may be structurally linked to the downstream interactions with NTS, possibly in a synchronized ratcheting manner conducive to polymerase translocation.

The work presented in this chapter was reproduced with permission from Elsevier Publishing:

Barnes CO, Calero M, Malik I, Graham B , Saphr H, Cohen A, Guowu L, Brown I, Zhang Q, Pullara F, Trakeselis M, Kaplan C, and Calero G. (2015) Crystal structure of a transcribing RNA Polymerase II complex reveals a complete transcription bubble. *Molecular Cell*, ***59****(2),* 258-269. doi: 10.1016/j.molcel.2015.06.034

3.2 EXPERIMENTAL METHODS

3.2.1 Purification of *S. cerevisiae* Pol II

Ten-subunit *S. cerevisiae* Pol II Δ*rpb4* deletion strain harboring a protein A-tagged Rpb1 was purified as previously described with minor modifications (*37*). Lysate was precipitated with 0.2% polyethylenimine and clarified at 30,000*g*. A saturated ammonium sulfate solution was added to the clarified supernatant to a final concentration of 60% and allowed to equilibrate at 4 ˚C for 3 h before centrifugation at 30,000*g*. The resulting pellet was dissolved in a 10-fold weight:volume ratio of resuspension buffer: 1x TEZ (50 mM Tris pH 7.5, 10 µM ZnCl, 0.5 mM EDTA), 2 mM DTT, 10% glycerol, and 1x protease inhibtors (100x stock: 2.5 mg Leupeptin, 5 mg Pepstatin A, 1.6 g Benzamidine, 0.85 g PMSF resuspended in 50 mL dry Ethanol) and adsorbed onto IgG resin. Pol II was eluted with overnight TEV cleavage at 4 ˚C, and further purified by heparin and size exclusion chromatography using well-established protocols (*44*). Reconstitution of 12-subunit Pol II was achieved using recombinantly expressed Rpb4/7 subunits as previously described (*42*).

3.2.2 Pol II-TIC-Tfg2 and Δ4/7 Pol II-TIC-TFIIF assembly

Tfg2 was purified as previously described (*100*). To assemble a Pol II transcribing complex (Figure 7), synthetic oligonucleotides, shown in Table 1, containing a single stretch of non-complementary bases and a 9-mer RNA were annealed above their melting temperatures and allowed to cool slowly to 4º C. The resulting nucleic acid scaffolds (NAS) were mixed with Pol II (3:1 molar ratio) and excess scaffold was removed using size exclusion chromatography

(Superdex200, GE LifeSciences) against Buffer A (25 mM Hepes pH 7.5, 100 mM KCl, 5 mM DTT, 0.5 mM EDTA, 10 µM ZnCl2). A 5-molar excess of Tfg2 was mixed with the Pol II-TIC in the presence of 0.2% Zwittergent 3-10, and dialyzed overnight at 4º C against Buffer A. Removal of excess Tfg2 was carried out using a second size exclusion step (Superdex200, GE LifeSciences) against Buffer A. An SDS-PAGE of the final complex is illustrated in Figure 7B.

Table 1. Synthetic nucleic acid sequences for transcribing complexes

Oligonucleotide	Sequence (5' → 3')
	NAS Sequences used for structural studies
9mer RNA	UCGAGAGGA
Scaffold 1 TS^	CCTACCGATAAGCAGACGATCCTCTCGAACCACGGACTCTTTATATACAAGCG
Scaffold 1 NTS	CGCTTGTATATAAAGAGTCCGTGGAAGCTCTCCTAGCAGTGCTTATCGGTAGG
Scaffold 2 TS^	CCTACCGATAAGCAGACGATCCTCTCGAACCACGGACTCTTTATATACAAGCGCGC
Scaffold 2 NTS	GCGCGCTTGTATATAAAGAGTCCGTGGAAGCTCTCCTAGGTCTGCTTATCGGTAGG
Scaffold 3 TS^	CTACCGATAAGCAGACGATCCTCTCGATGCATTGACTCATCGACG
Scaffold 3 NTS	CGTCGATGAGTCAATCGTAGCTCTCCTAGCACTGCTTATCGGTAG
	DNA Sequences for 2-AP Experiments
9mer RNA	UCGAGAGGA
Comp NTS#	GCCATAAAGATCCGTGCTTCGAGAGGATTTTTGCTTATCGGTACG
Non-Comp NTS# (i+3)	GCCATAAAGATCCGTGGAAGCTCTCCTAATTTGCTTATCGGTACG
Non-Comp NTS# (i+5)	GCCATAAAGATCCGTGGAAGCTCTCCTAAAATGCTTATCGGTACG
i2TS*	CGTACCGATAAGCAAAA**A^{P}**ATCCTCTCGAAGCACGGATCTTTATGGC
i3TS*	CGTACCGATAAGCAA**A^{P}**AATCCTCTCGAAGCACGGATCTTTATGGC
i5TS*	CGTACCGATAAGC**A^{P}**AAAATCCTCTCGAAGCACGGATCTTTATGGC
i8TS*	CGTACCGATA**A^{P}**GCAAAAATCCTCTCGAAGCACGGATCTTTATGGC

^Nucleic acid scaffold (NAS) template strands used for crystallization where the underlined region denotes the non-complementary stretch of base pairs to make an artificial bubble, and A represents the i+1 site

#Non-template DNA strands used for 2 – aminopurine experiments, where compNTS represents the fully complementary strand, and non-comp NTS represents bubble DNA.

*Template strand DNA where A^{P} represents 2 – aminopurine, and A designates the i+1 site

Purification of TFIIF and assembly of Δ4/7 Pol II-TFIIF complex was achieved similarly to Pol II – Tfg2 transcribing complexes (Figure 7), with a more detailed protocol found in our previous publication (*101*). An SDS-PAGE of the final complex is illustrated in Figure 7C. Ethidium bromide staining of sample confirmed the presence of the NAS (Figure 7C, red arrow).

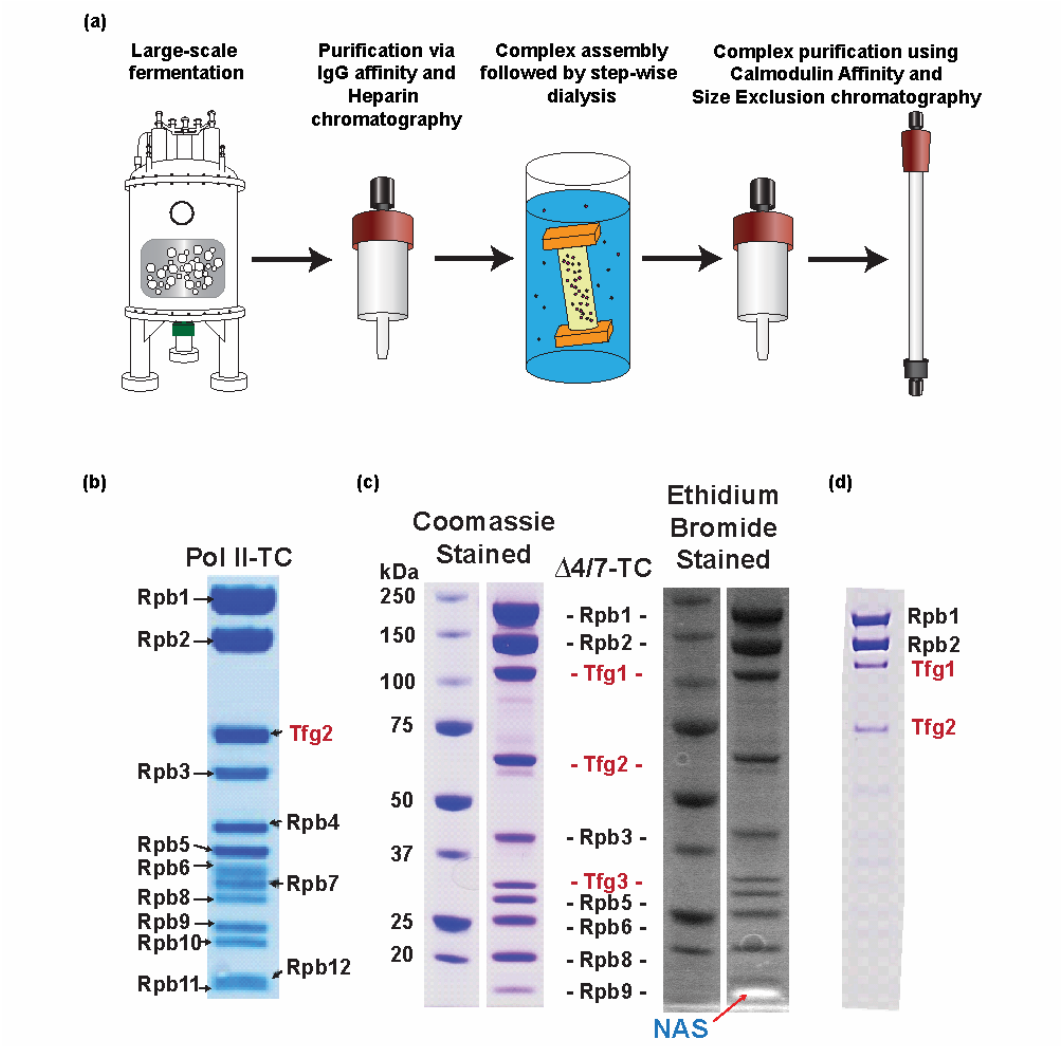

Figure 7. Complex purification and assembly.

(a) *Schematic flow-chart of steps taken for protein purification and complex assembly.* ***(b)*** *SDS-PAGE of Pol II-Tfg2 complex.* ***(c)*** *SDS-PAGE of Δ4/7 Pol II TFIIF complex, also depicting NAS after ethidium bromide staining (red arrow).* ***(d)*** *Crystals of Δ4/7 Pol II-TFIIF complex were washed with reservoir solution and 8X loading buffer was added to the drop containing crystals and subjected to SDS-PAGE. The four largest proteins are easily discernible from the gel confirming the presence of the complex in the crystals.*

3.2.3 2 – aminopurine fluorescence spectroscopy

Experiments presented here were performed in tandem with Dr. Brian Graham from the lab of Dr. Michael Trakeselis. DNA and RNA oligonucleotides are listed in Table 1. DNA/RNA hybrids of the TS and RNA were made by annealing oligonucleotides with a 1:2.5 molar ratio. Excess RNA was removed using a Superdex75 10/300 column (GE Lifesciences) against Buffer A, and nucleic acid scaffolds were assembled by annealing equimolar concentrations of the TS/RNA hybrids with NTS. To assemble transcribing complexes, 350 nM RNA Polymerase II was incubated with 300 nM NAS for 1 h at room temperature in Buffer B: 50 mM Hepes (pH 7.0), 120 mM NaCl, 5 mM βME, 2 mM MgCl2, 0.1 mM EDTA and 1 μM ZnCl2. Steady-state fluorescence measurements were acquired in buffer B using a Fluoromax-3 (HORIBA Scientific). Excitation wavelengths included both 280 and 315 nm and fluorescence emission was collected from 340-400 nm. The entrance and exit slits were adjusted for optimal intensities. Spectra were collected for buffer, DNA and polymerase alone to correct for background fluorescence as described elsewhere (*102*). All measurements were performed at 25 ºC in triplicate.

3.2.4 Crystallization and Refinement

Several conditions were used for obtaining crystals of Pol II – TICs (Table 2). Scaffold 1 (i+5) was the primary NAS used for structure determination and complexes were verified by SDS-PAGE analysis of crystals (Figure 7D). The structures were solved by molecular replacement (MR) using 12-subunit Pol II PDB:ID 3FKI (*103*) for Pol II-TICs and Pol II-TIC-Tfg2 and 10-subunit Pol II PDB:ID 1R9T (*43*) for Δ4/7 Pol-II-TIC-TFIIF as search models, respectively, in

Molrep (*104*). MR models were refined using the program Buster (*105*), Refmac (*106*, *107*) and CNS (*108*) followed by several cycles of manual building with B-factor sharpening in COOT (*109*, *110*). All figures were rendered using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

Table 2. Crystallization conditions for Pol II - TICs

Complex	Crystallization condition
Pol-II-TIC-i+5	6% PEG 6000, 0.2 M ammonium acetate, 0.3 M sodium acetate, 0.1 M Hepes pH 7.0
Pol-II-TIC-Tfg2-i+5	10% PEG 4000, 75 mM NaCitrate, 0.1 M MES pH 6.5
Pol-II-TIC-Tfg2-i+2	12% Peg 4000, 75 mM potassium thiocyanate, 0.1 M MES, pH 6.5
Pol-II-TIC-i+3	6-8% Peg 6000, 0.2 M ammonium acetate, 0.3 M sodium acetate, 0.1 M Hepes pH 7.0
Δ4/7 *Pol-II-TIC-TFIIF-i+5*	8% Peg 4000, 75 mM sodium malonate, 0.1 M Hepes pH 7.0

^ All conditions include 10 mM DTT, and where cryo-protected into mother liquor + 32% glycerol.

3.3 RESULTS

3.3.1 Design, Assembly, and Crystallization of a Pol II-TIC

Assembly of a *S. cerevisiae* Pol II-transcribing intermediate complex (Pol II-TIC) was achieved by mixing Pol II with pre-assembled nucleic acid scaffolds (NAS, Table 1). The main scaffold used for our experiments (scaffold 1) consisted of two synthetic DNA oligonucleotides (53-nucleotides long), featuring upstream and downstream duplexes, a non-complementary stretch of 15 nucleotides to generate a synthetic transcription bubble, and a 9-mer RNA complementary to the template strand to form a DNA-RNA hybrid (Figure 8A). Structural analysis presented in this paper utilized scaffold 1 within the TIC. The use of a non-complementary NTS to generate a TB is analogous to use of chain-terminators or non-hydrolyzable substrates to examine Pol II-substrate interactions or interpretation of Pol II inhibitor-bound complexes as translocation intermediates. Our current structure interpreted in similar light still allows for a number of valuable insights. The number of non-complementary bases used in the design of the TB was based on crystal structures of partial Pol II transcribing complexes including PDB:ID 1Y1W (*45*) and PDB:ID 2NVZ (*49*). These structures show at the downstream end, base complementarity at positions i+3 and i+5, respectively (where i+1 indicates the nucleotide addition site and i-1 the first base of the nascent RNA transcript), and at the upstream end, a partial template strand reaches position i-9 below arch residues (comprising rudder ($Rpb1^{312-319}$) and fork loop 1 (FL1, $Rpb2^{470-480}$)) (Figure 8B). However, steric clashes with arch residues at this position suggested that at least two additional nucleotides are required to allow template and non-template strand annealing. Collectively, these observations suggested an artificial bubble size with a minimum of 14 nucleotides for *in vitro* structural studies of a transcribing Pol II.

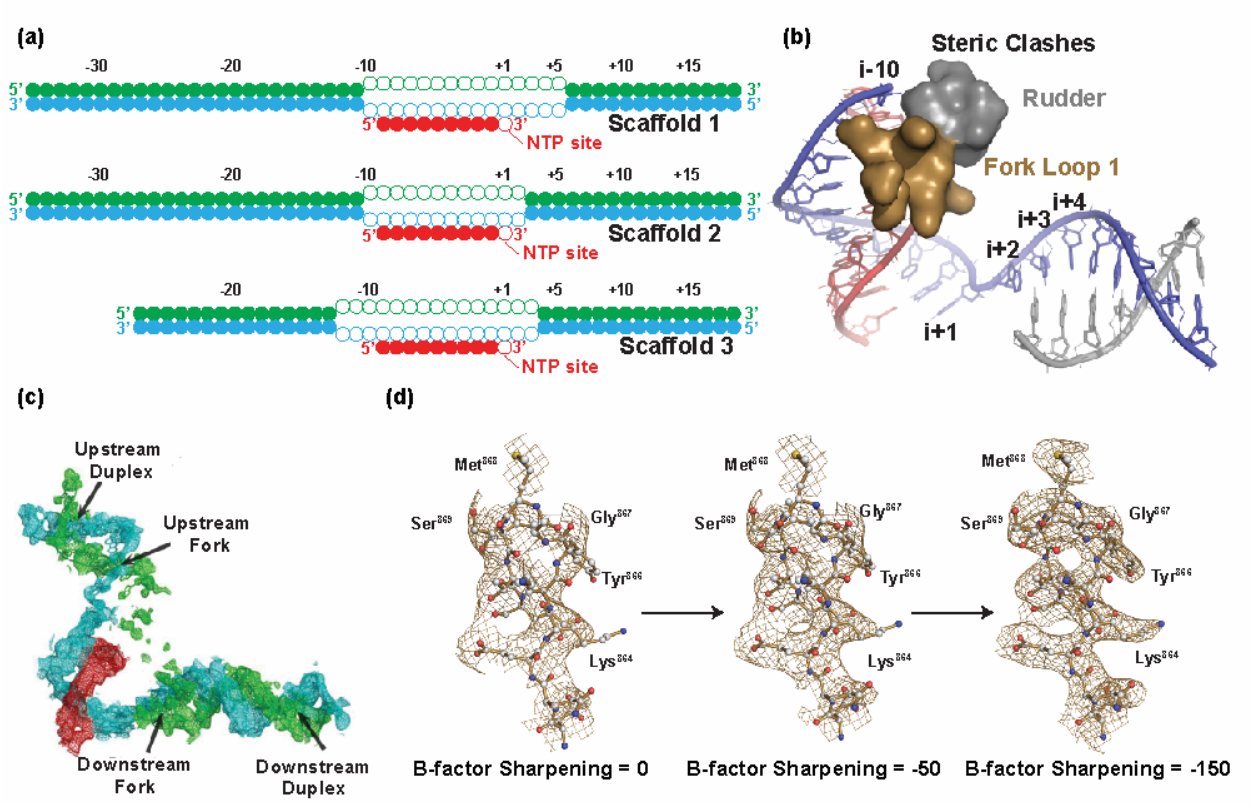

Figure 8. Design of Pol II transcribing complex.

(a) *Schematic representation of scaffolds used in our structural studies, where the open circles represent the non-complementary region to generate artificial transcription bubbles (template strand: cyan, non-template strand: green, RNA: red).* ***(b)*** *A conservative model of the size of the bubble could be estimated using PDB:ID 1Y1W), where complementarity was observed at i+3. Upstream clashes with the "arch" would require 2 additional bases before annealing could occur.* ***(c)*** *Electron density map $2F_{obs}$-F_{calc} contoured at 0.8σ of a crystal comprising Pol II + scaffold 1 in the absence of Tfg2 or TFIIF. Partial electron density is observed for the upstream duplex and for the downstream fork and duplex, however no density is observed for the non-template strand within the transcription bubble (TS: cyan, NTS: green RNA: red).* ***(d)*** *Example of B-factor sharpening on Pol II residues at the Rpb2 wedge, contoured at 1.0σ. Application of negative B-factors results in increased clarity for high resolution features, especially for large side chains (ie Arg, Lys, Phe, Trp, Tyr) at a relatively low resolution* (*111*).

Initially, crystals of Pol II bound to scaffold 1 (15 nucleotide TB) showed weak electron density for the upstream duplex but none for the NTS (Figure 8C). In search for factors that could contribute to a stabilized transcription bubble, we assembled Pol II or 10-subunit Pol II (lacking Rpb4 and Rpb7 subunits, Δ4/7) transcribing complexes with TFIIF or its 45 kDa β-subunit, Tfg2 (Figure 7B,C). Two sets of crystals were obtained using PEG 4000 and low salt (Table 2). The first transcribing complex comprises Δ4/7-TFIIF (heretofore referred as Δ4/7-TC for simplicity) and the second comprises Pol II-Tfg2 (heretofore referred as Pol II-TC). Structures were solved by molecular replacement (MR) using Δ4/7 or Pol II as search models (see **section 3.2.5**). An initial unbiased F_{obs}-F_{calc} map revealed the presence of extra density corresponding to upstream dsDNA, the NTS (Figure 9A) and three previously disordered regions of Rpb2 (Figure 13A,B). Unfortunately, any additional density for Tfg2 or TFIIF was non-interpretable due to either high mobility or partial occupancy in the crystals. Nevertheless, the presence of TFIIF within the crystals (Figure 7D) was essential to reveal the full structure of the transcription bubble in our complexes, suggesting an important role in non-template strand stabilization.

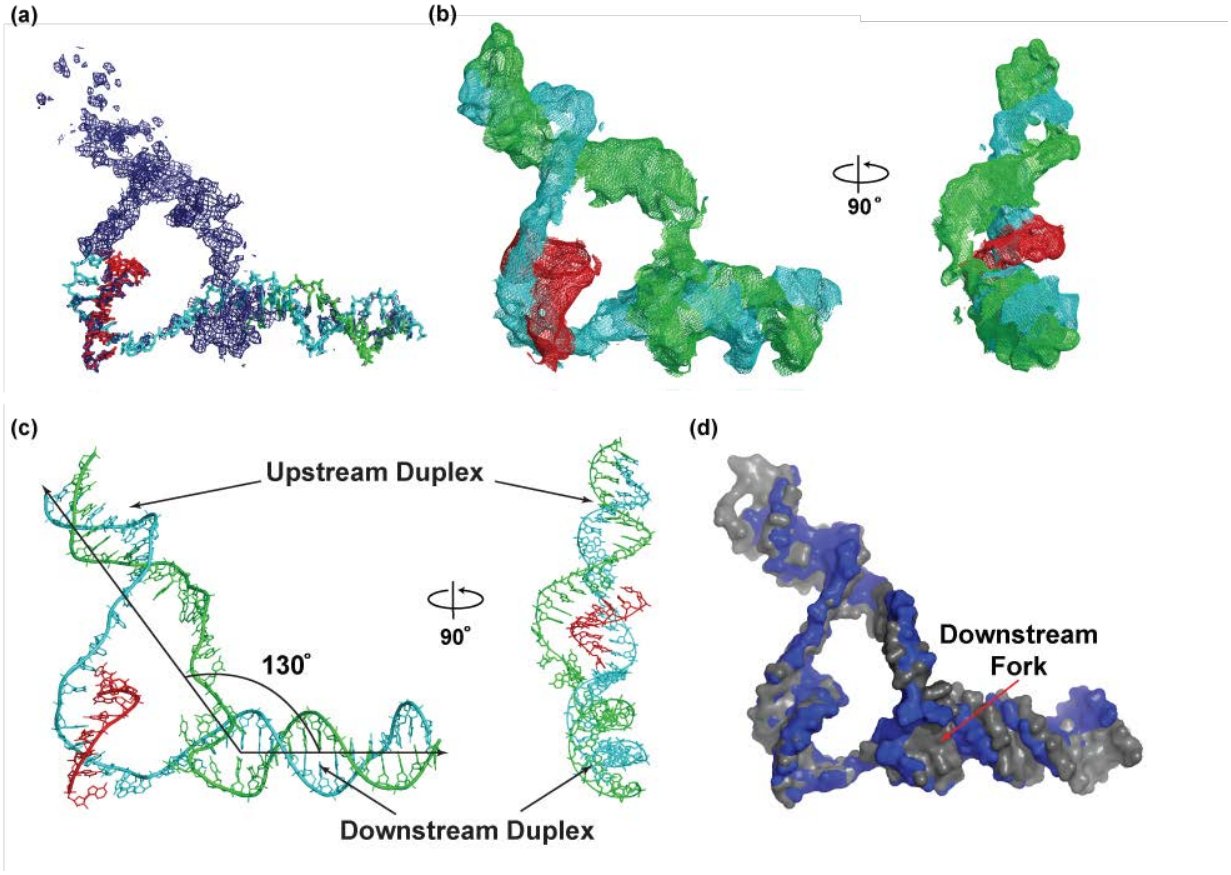

Figure 9. Architecture of the complete nucleic acid scaffold.

a) Difference Fobs-Fcalc electron density map contoured at 2σ. The following color scheme for the NAS will be used throughout: cyan, template strand (TS); green, non-template strand (NTS); red, RNA transcript. b) Front and side views of the Fobs-Fcalc map for the final refined map contoured at 1.0σ. c) Cartoon representation of the 38-nucleotide refined nucleic acid scaffold (NAS); downstream and upstream duplexes form an angle of approximately 130º degrees. d) Surface representation of the overlay between Δ4/7-TC (blue) and Pol II-TC (grey). The two structures overlay remarkably well with minor structural differences occur at the downstream fork.

The MR model was refined using the programs Buster (*105*), Refmac (*106*) and manual building with B-factor sharpening in COOT (see Table 3) (*109, 110*). The NAS was built into the electron density using the characteristic features of the DNA-RNA hybrid as register (Figure 9A). The final refined $2F_{obs}$-F_{calc} map for Δ4/7-TC is illustrated in Figure 9B. The full observable DNA scaffold (38 nucleotides long) spans the length of Pol II and comprises: the downstream duplex, the DNA-RNA hybrid and two previously uncharacterized regions, an upstream duplex and the full transcription bubble including the non-template strand (Figure 9B,C). Overlay between the DNA scaffolds from Δ4/7-TC and Pol II-TC, show minor differences mainly located at the downstream fork (Figure 9D). Pol II regions involved in DNA binding include: 1) The previously described Rpb5 "jaw" and Rpb1 "clamp" residues that interact with the downstream duplex (*38*); 2) Rpb2 "wedge" residues ($Rpb2^{862\text{-}874}$) that interact with the upstream duplex; and 3) "arch" residues that interact with the upstream fork (Figure 10A,B). For clarity, the presentation of our structural findings is based on the 10-subunit Pol II (Δ4/7) transcribing complex comprising TFIIF and scaffold 1, which has been labeled as Δ4/7-TC unless otherwise noted.

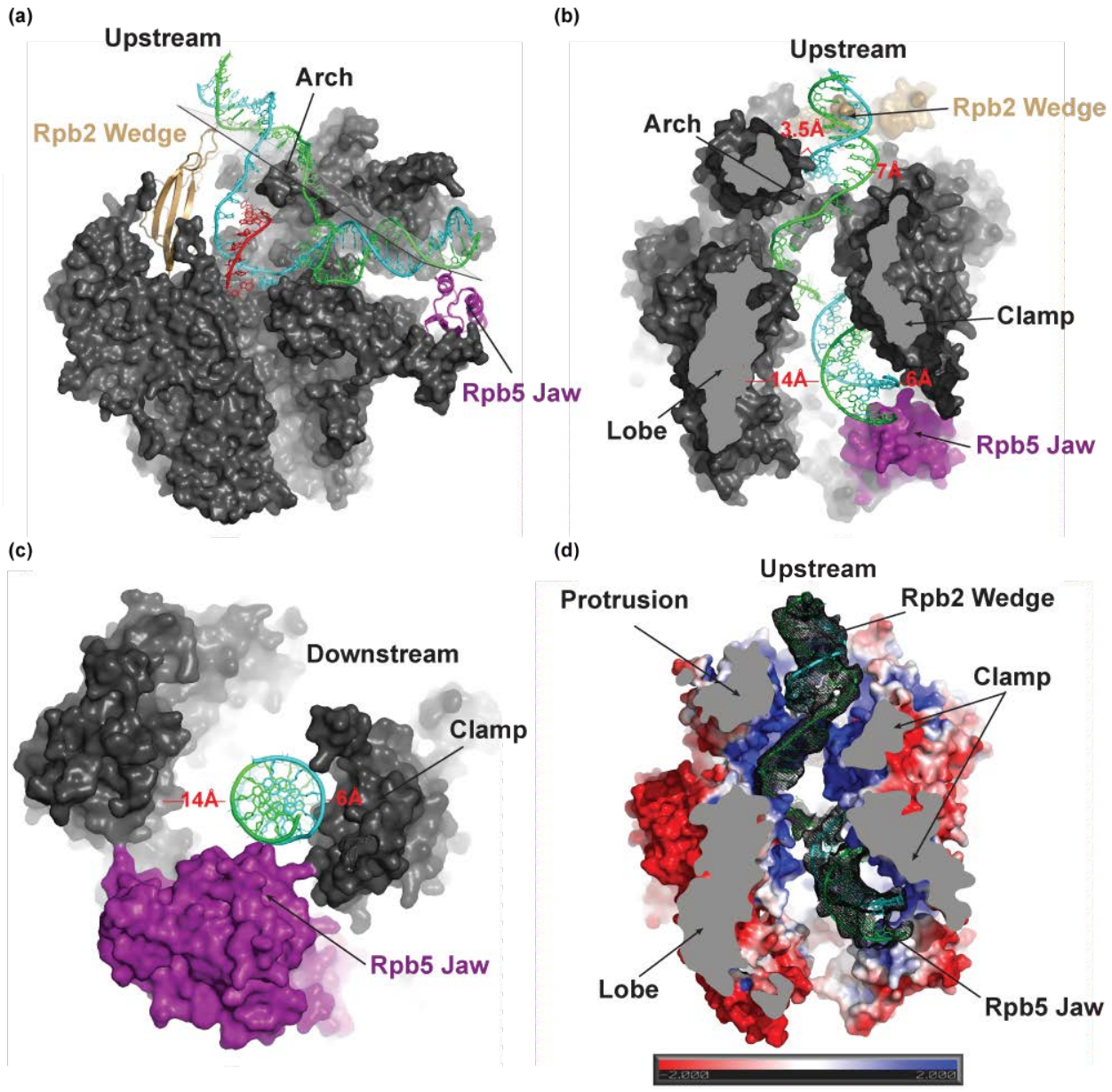

Figure 10. Overall structure of Pol II transcribing intermediate complexes.

(a) *Surface representation (side view, Rpb2 removed) illustrating the position of the NAS inside Pol II. Wedge, jaw and arch interactions with the NAS lie on almost a perfect plane, possibly to minimize strain during elongation.* ***(b,c)*** *The NAS binds asymmetrically inside Pol II's cleft, more prominently at the downstream end where observed clamp-DNA distances of ~6 Å vs. lobe DNA distances of ~14 Å are due to interactions with clamp and jaw residues.* ***(d)*** *Surface electrostatic representation calculated using the APBS (112) suite in PyMOL to illustrate how the NTS follows a path of positively charged residues inside Pol II's cleft (lobe and protrusion).*

Table 3. Crystallographic Data and Refinement Statistics

	Pol II-Scaffold 1 (i+5)[a] (APS-GM/CA)	*Δ4/7 –TC (i+5)[a] (APS-GM/CA)*	*Pol II-TC (i+5)[a] (SSRL-11.2)*	*Pol II-Tfg2 – Scaffold 2 (i+2)[a] (SSRL-11.2)*	*Pol II-Scaffold 3 (i+3)[a] (APS-GM/CA)*
Data Collection[b]					
PDB:ID	5C44	5C4J	5C4X	5C4A	5C3E
Nucleic Acid	Scaffold 1	Scaffold 1	Scaffold 1	Scaffold 2	Scaffold 3
Space Group	C2221	C21	C2221	C2221	C2221
Unit cell (Å)	220.2 391.8 282.3	280.7 223.3 156.4	220.7 393.3 281.6	219.8 396.7 273.6	219 390.9 278
α, β, γ (°)	90 90 90	90 98.1 90	90 90 90	90 90 90	90 90 90
Wavelength (Å)	1.03	0.979	0.979	0.978	1.03
Resolution (Å)[c]	120-3.9	174-4	50-4	200-4.2	178-3.7
Unique Reflections	106,511	80,485	103,153	80,062	126,815
Completeness (%)	96 (93.8)	97(96.1)	97.3(96.4)	99(98.3)	94.88(78)
Redundancy	4.3 (3.2)	5.1(4.3)	3.8(3.4)	4.5(3.7)	3.6(2.8)
<I/σI>	8.2(1.2)	10.3(3.7)	12.5(1.95)	14.2(2.1)	8.5(1.1)
Mosaicity (°)	0.6	0.6	0.35	0.55	0.35
R_{merge} (%)	13.8(49)	14.7(61)	9.1(42)	8.5/(47)	12.5(56)
Refinement					
No. Atoms	33086	30259	33673	32281	33086
R_{cryst}/R_{free} (%)	22.3/25.4	21/26	21.5/23.2	23.53/27.47	23.2/27.62
Refinement Program	Buster/CNS	Refmac/CNS	Buster/CNS	Refmac/CNS	Refmac/CNS

[a]Indicates the position of complementary base pairing in the downstream bubble.
[b]Numbers in parentheses correspond to the highest resolution shell
[c]Resolution limits were extended to include weak intensity data (*113*). Using the traditional criterion of I/σI > 2.0, resolution limits are 4.15 Å and 3.9 Å for Pol II-Scaffold 1 and Pol II-Scaffold 3 complexes, respectively.

3.3.2 Pol II interacts with the minor groove of upstream and downstream DNA duplexes using two domains located 90 Å apart.

Our refined models show that Pol II interacts with the upstream duplex, which appears as if propped by a hairpin loop or "wedge" ($Rpb2^{862-874}$) –a vertical extension of wall residues ($Rpb2^{855-861}$)– that engages the minor groove of the double helix (Figure 10A). At the tip of the wedge, Met^{868} lies between template and non-template strands, while the amide backbone of

Gly^{867} appears to form hydrogen bonds (H-bond) with two contiguous phosphates on the non-template strand (Figure 11A). No crystal contacts that could potentially stabilize the conformation of wedge residues or the upstream duplex were present in the two different crystal forms (Figure 12A,B). The framework of the wedge –a hairpin loop within a long concave five-strand β-sheet– is conserved in all multi subunit DNA-directed RNA polymerases (Figure 11B). Moreover, overlay of the human mitochondrial polymerase elongation complex (mtRNAP) (*114*) and our structure shows that the upstream duplex and fork adopt similar conformations (Figure 11C). This is particularly interesting given the scarcity of conserved structural elements between the two structures (Figure 12C).

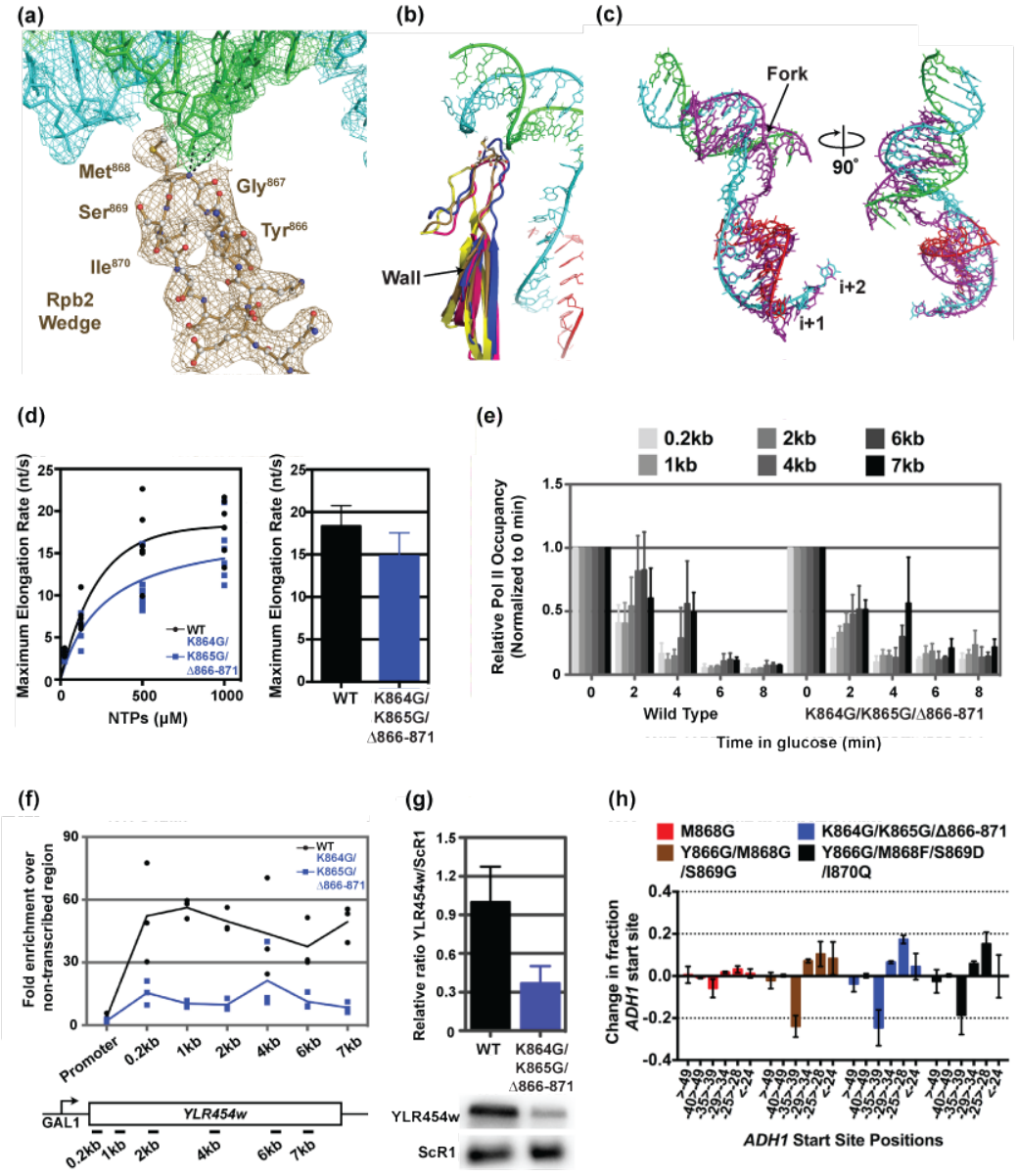

Figure 11. Rpb2 Wedge Residues: Structure, Conservation, and Function (see also Appendix A)

(a) *Interaction of wedge residues with the minor groove of the upstream duplex. A refined* $2F_{obs}$ *-* F_{calc} *map contoured at 1.0* σ *is also illustrated.* ***(b)*** *Conservation of wedge structure: yellow, T. thermophilus* (*47*)*; hot pink, archaea (S. sulfactarius)* (*115*)*; sand, S. cerevisiae; blue, S. pombe* (*116*)*; and purple, 14-subunit Pol I (S. cerevisiae)* (*117*)*.* ***(c)*** *Overlay of mtRNAP-upstream duplex (PDB: 4BOC) with our structure about template strand i+1 and i+2.* ***(d)*** *In vitro elongation rate of WT Pol II and an rpb2 wedge deletion mutant (K864G/K865G/D866-871). Elongation rate determined on nucleic acid scaffolds at a number of NTP concentrations followed by non-*

linear regression of the rates for determination of maximum elongation rates (bar graph, error bars indicate range of 95% confidence interval). ***e)*** *In vivo apparent elongation rates for WT Pol II and the rpb2 wedge deletion mutant at a galactose-inducible reporter gene determined by ChIP upon glucose shutoff of transcription (schematic of reporter in panel f). Values are normalized to 0 min of glucose and error bars represent SD of the mean for three independent experiments.* ***f)*** *Steady state occupancy for wt Pol II and the rpb2 wedge deletion mutant at a galactose inducible reporter gene under galactose induction determined by chromatin IP (schematic of reporter with positions of PCR amplicons shown below) (n=3 independent experiments).* ***g)*** *Steady state RNA levels of reporter used in* ***F*** *for wt Pol II and the rpb2 wedge deletion mutant. Values were normalized to SCR1 levels (a Pol III transcript) and averaged (n=3) with error bars representing standard deviation.* ***h)*** *Primer extension analysis for ADH1 transcripts in rpb2 wedge alleles (left) with quantification on right showing average change in fraction of ADH1 starts in various positions relative to wild type. Error bars represent standard deviation of the mean (n=3).*

To assess a role of Rpb2 wedge residues during *in vivo* transcription, a number of alleles were constructed and characterized for growth phenotypes consistent with transcription defects in collaboration with the lab of Dr. Craig Kaplan (see **Appendix A**, Figure 12D, left panel). We observed sensitivity to mycophenolic acid (MPA), which can be indicative of *IMD2* transcriptional phenotypes (*118*). Indeed, we found that Rpb2 wedge alleles were generally defective for induction of *IMD2* gene expression in the presence of MPA (Figure 12D, right panel). In order to more directly assess the role of the Rpb2 wedge, the lab of Dr. Kaplan examined the K864G/K865G/Δ866-871 *rpb2* wedge allele for *in vitro* or *in vivo* elongation phenotypes (see **Appendix A**, Figure 11D,E). We found that this particular *rpb2* wedge allele did not confer robust elongation defects *in vitro* or *in vivo*, though it did confer very strong defects in steady state Pol II reporter gene occupancy and expression (Figure 11F,G). The defects observed are consistent with strong defects in initiation, and we observed altered transcription start site selection consistent with altered initiation *in vivo* (Figure 11H) (*118*).

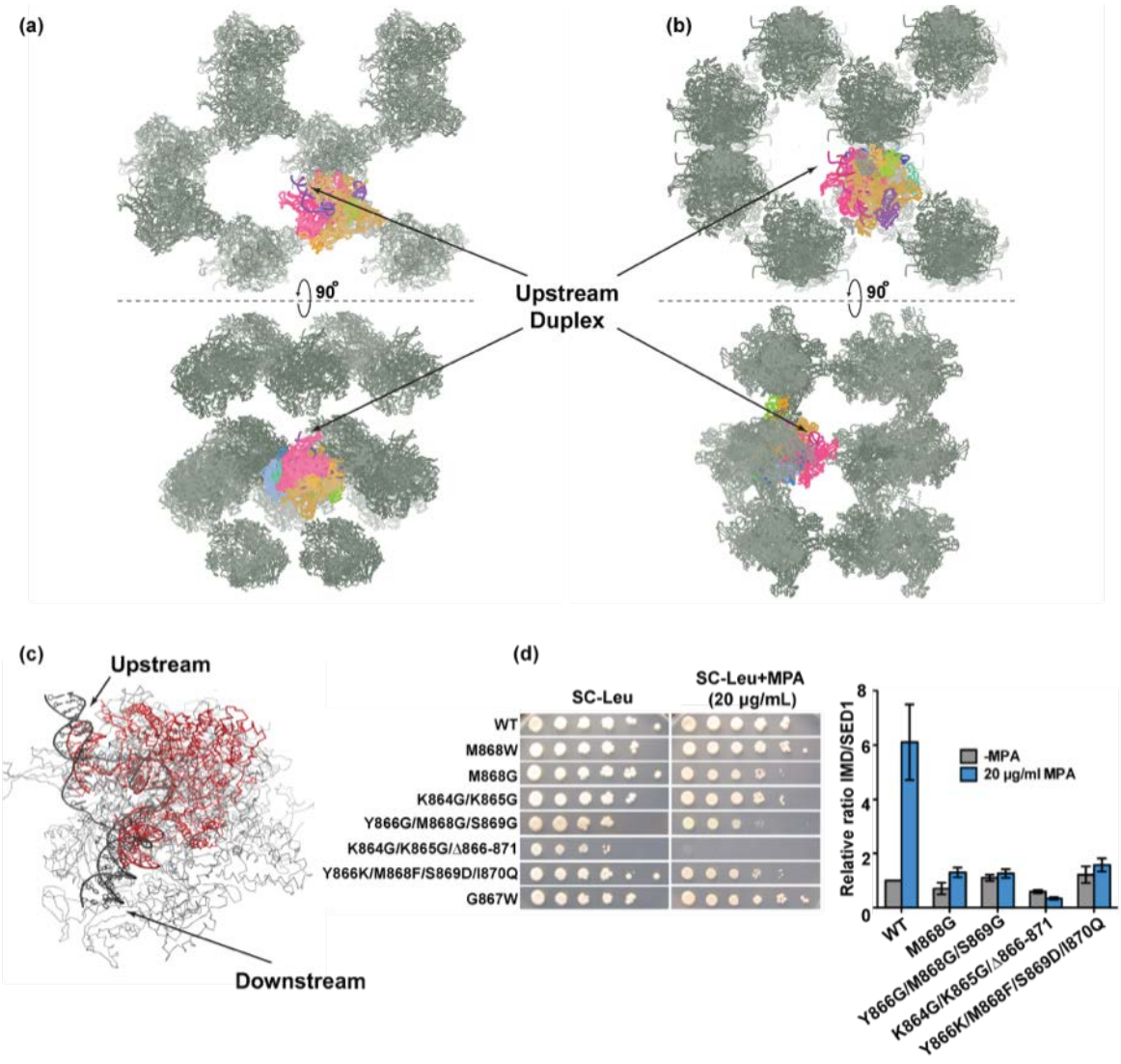

Figure 12. Pol II –TIC crystal contacts and the role of Rpb2 wedge residues in upstream duplex stabilization (see also Appendix A)

(a,b) *Crystal symmetry of Δ4/7-TC crystallized in C21 space group and Pol II-TC in space group C2221. Symmetry molecules were visualized in PyMol with upstream DNA positions highlighted (black arrow). The upstream duplex is solvent exposed indicating that the DNA architecture was not an artifact of crystal packing.* ***(c)*** *Ribbon and cartoon representation of the overlay between the elongating human mitochondrial RNAP (red) PDB:ID 4BOC* (*114*) *and our structure (grey) about positions i+1 and i+2 on the template strand.* ***(d)*** *Left panel – Serial dilutions of yeast strains containing Rpb2 wedge alleles (mutations as designated on figure) on various media. Right panel – Quantitative analysis of Northern blotting for IMD gene expression (normalized to SED1 as loading control) in presence or absence of 20µg/ml mycophenolic acid (MPA). Error bars are standard deviation of the mean (n=3).*

Pol II contacts with the downstream duplex encompass Rpb5 (jaw) and Rpb1 (clamp-head) domains (Figure 10B,C). Specifically, Rpb5 jaw residue Pro^{118} (from helix 118-124) is positioned inside the minor groove of the DNA double helix and Thr^{117} (from loop 112-117) and Ser^{119} locate within H-bond distance to non-template strand positions i+15 and i+16 (Figure 13, also observed in PDB:IDs 1R9T, 2NVQ and 2NVZ). Rpb1 clamp-head residues Lys^{100}, Lys^{101}, Lys^{143}, and Arg^{175} locate within H-bond distance of the phosphate chain of non-template strand positions i+8 to i+10 (Figure 13D). As a result of these interactions, the downstream duplex is asymmetrically positioned inside the cleft (Figure 10B,C). Interestingly, contacts resembling clamp-head interactions in *T. thermophilus* polymerase, are located on a helix-loop-helix domain in its β' subunit (Figure 13E) forming H-bonds with non-template strand i+8 to i+9. However, interactions with jaw residues are observed only in archaeal and eukaryotic polymerases, since bacterial polymerases lack Rpb5 homologues.

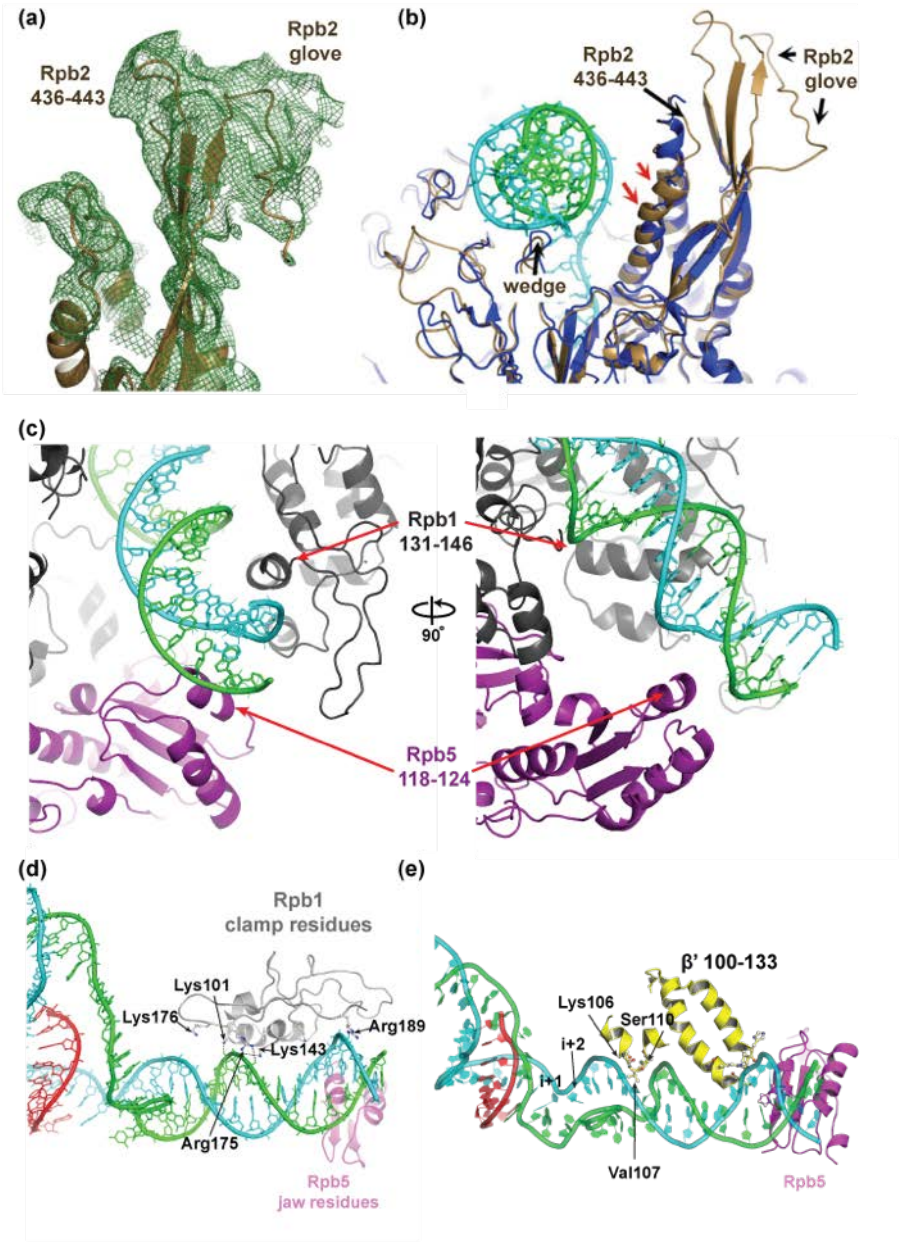

Figure 13. Pol II contacts with upstream and downstream duplexes.

a) *$2F_{obs}$-F_{calc} electron density map contoured at 1.0σ to illustrate the presence of traceable electron density for previously disordered Rpb2 regions.* ***b)*** *Cartoon representation of the overlay of our structure with PDB:ID 1Y1W (illustrated in sand and blue, respectively); template and non-template strands are illustrated in cyan and green respectively. Presence of TFIIF or Tfg2 allowed tracing of two previously unfolded regions of Rpb2, the first one folds into a 3-strand beta sheet that resembles a catcher's glove (Rpb2 glove) and includes residues 71-90 and 131-164. The second one includes loop 436-443 from Rpb2 protrusion domain. Overlay between our structure and PDB:ID 1Y1W illustrate conformational changes in helix 443-465 (red arrows) possibly elicited by the close*

proximity of the upstream duplex. ***c)*** *Cartoon representation of Pol II stabilizing the downstream duplex illustrates that Rpb1 (silver) and Rpb5 (magenta) helices "thread" the major and minor grooves of the DNA respectively.* ***d)*** *Cartoon and ball & stick representation of Pol II interactions with the downstream duplex, where positively charged residues of Rpb1 clamp domain (silver) contact the i+7 to i+9 phosphates of the non-template strand. e) Contacts resembling head-clamp interactions in T. thermophilus polymerase (PDB:ID 2PPB) are located on a helix-loop-helix domain in the β′ subunit. These include: 1) Val107 which is inserted in the minor groove of the double helix, 2) the backbone nitrogen of Ser110 interacting with i+8 phosphate of the non-template strand and 3) Lys106 interacting with phosphate template strand i+4.*

3.3.3 Strand annealing and 3-way coordination of TS, NTS and RNA at the upstream fork

The transcription bubble in our complexes lies within upstream and downstream duplexes and was enforced by non-complementarity between the template and non-template strands at positions i+5 through i-10 (Figure 8A, scaffold 1). At the upstream "closing" end of the bubble, rudder and fork-loop 1 (FL1) residues come in close proximity to form an "arch" located 25-30 Å above the bridge helix (Figure 10A). The arch is situated in between the template and non-template strands, and physically marks the upstream boundary of the bubble (Figure 14). Arch residues adopt unique conformations that allow simultaneous coordination of the nucleic acid scaffold (Figure 14A). As the template strand separates from the RNA transcript (i-8) and emerges from Pol II's active site, non-specific packing interactions with FL1 residues and potential salt bridges between the phosphate chain and arch residues guide the template strand in a straight conformation towards its junction with the non-template strand (Figures 14A,B and 15A).

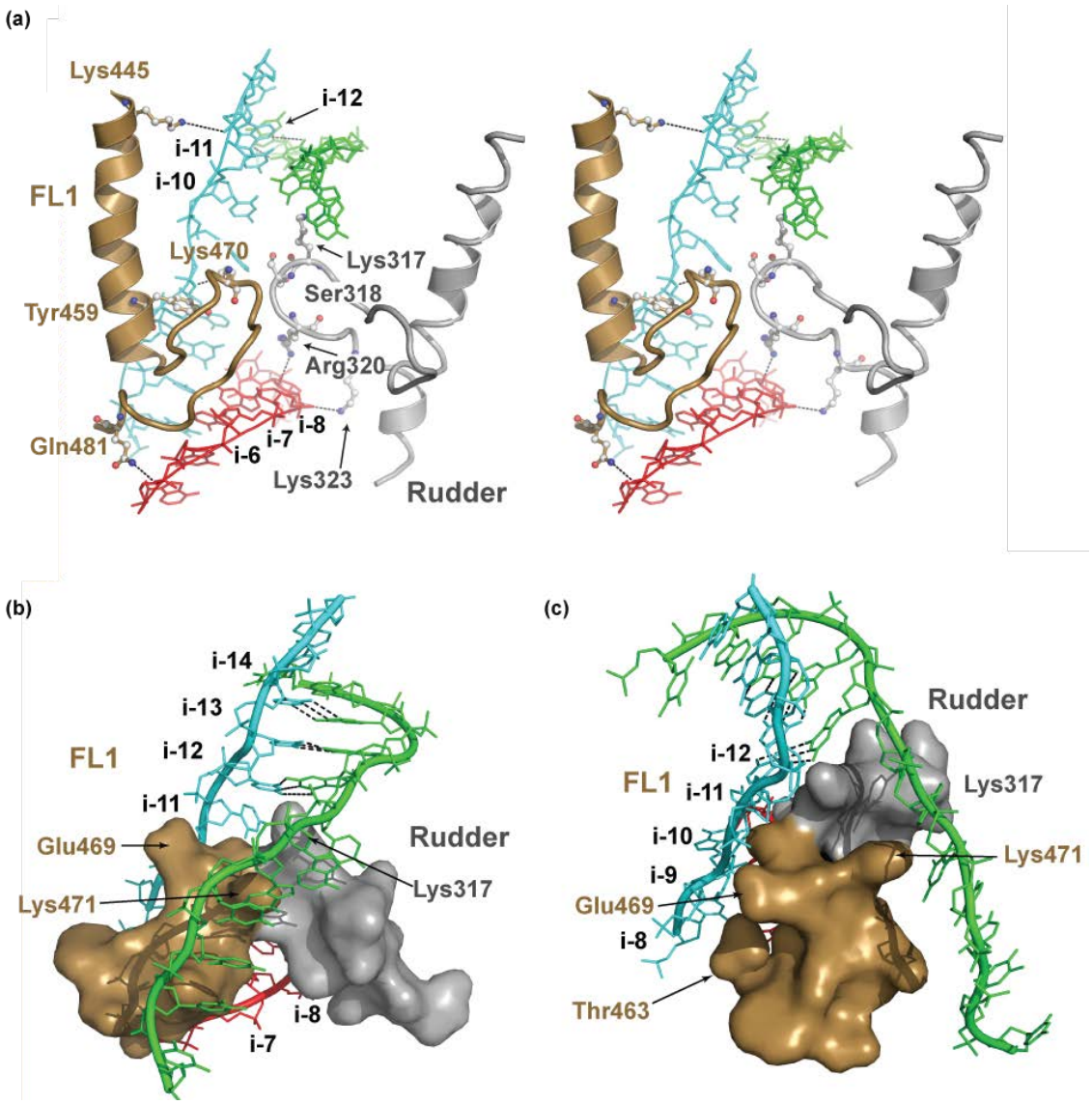

Figure 14. Architecture of the upstream fork junction

(a) *Stereo-view of the tripartite coordination of the three nucleic acid chains. Arch residues at the back, top and bottom adopt unique conformations – with respect to apo- and elongation structures – that interact with TS, NTS and the RNA transcript respectively. Rpb1 residue Arg320 forms a H-bond with the 2'-OH of the nascent transcript at position i-8.* ***(b,c)*** *Front view (b) and side view (c) of the upstream (closing) end of the TB. Rudder (silver) and FL1 (sand) residues reach within 4 Å across the midline to form an arch that provides a scaffold for TS and NTS annealing at i-12. Contacts include packing interactions between TS i-8 and Tyr459 and potential H-bonds between Rpb2 residues Thr463 with TS i-8, Glu469 with TS i-10 and Lys470 with TS i-9. Lys317 participates in contacts with NTS i-11 and i-12 (see also Figure 15)*

Once above the arch, template strand nucleotides anneal with the non-template strand at i-12 (Figure 14B,C). Stabilization of the nucleic acid scaffold by arch residues include, interactions between template strand i-8, i-9 and i-11 with FL1 residues, and between non-template strand i-11 and i-12 with rudder residue Lys^{317} (Figures 14 and 15B) (*119*). In addition to contacts that arch residues make with the template and non-template strands, rudder residue Arg^{320} reaches within H-bond distance of the 2´ hydroxyl of the 8^{th} RNA base of the nascent transcript (Figures 14A and 15B, observed also in PDB:ID 1Y1W). This interaction is conserved in bacterial polymerases, where rudder Arg^{598} forms a H-bond with the 2´ hydroxyl group of the 7^{th} base on the nascent transcript (Figure 15C, (*120*)) and has been corroborated by cross-linking experiments (*121*).

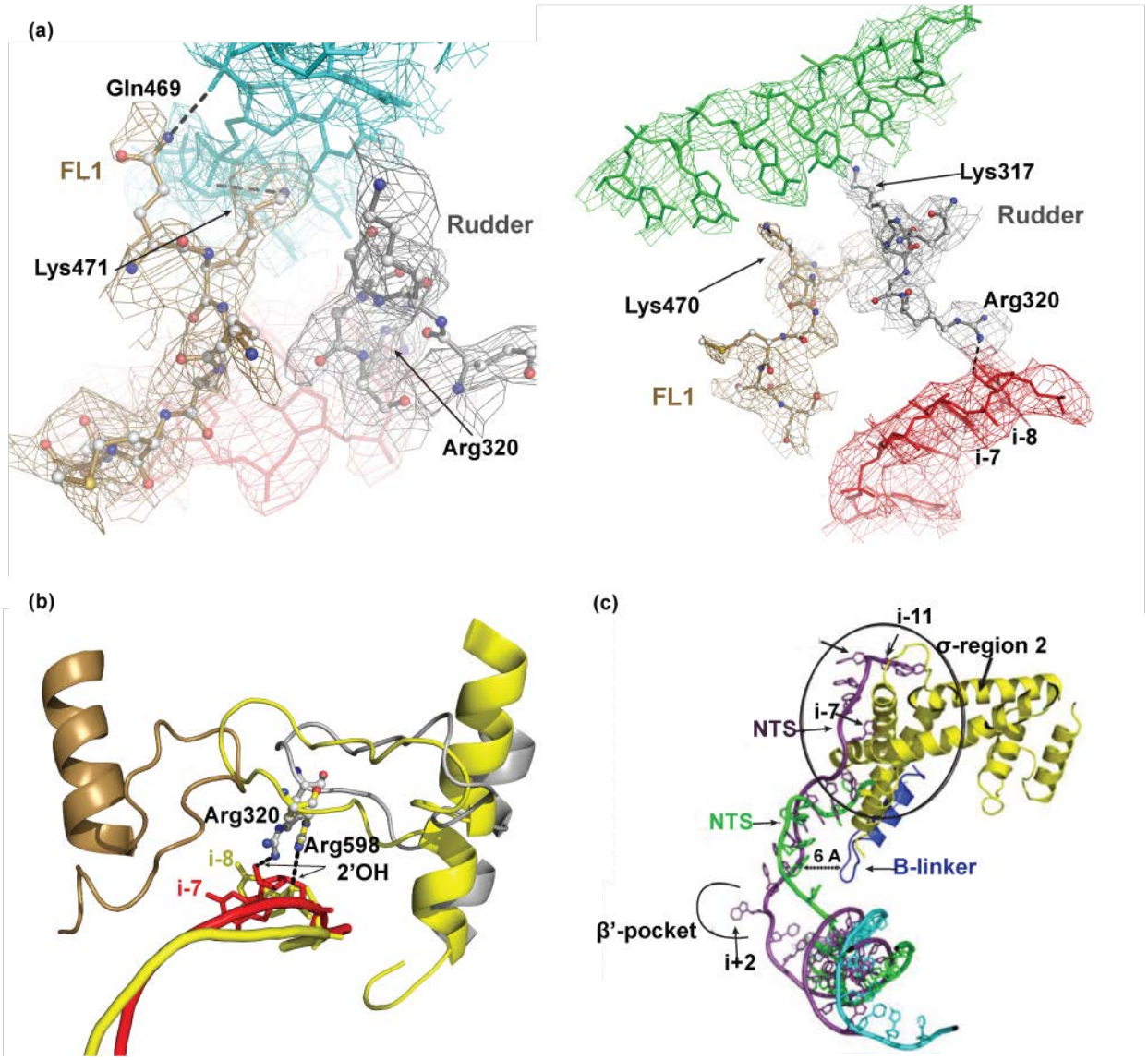

Figure 15. Comparison of eukaryotic and bacterial RNA polymerases

(a,b) *Architecture of the transcription bubble at the upstream (closing) end with* ***(a)*** *template strand (blue) and arch residues.* ***(b)*** *non-template strand (green), RNA (red) and arch residues. A refined 2Fobs-Fcalc electron density map (contoured at 1 σ) was calculated around both regions.* ***c)*** *Interactions between rudder residues Arg320 (S. cerevisiae, silver) or Arg598 (T. thermophilus, yellow) with the 2′ hydroxyl of the nascent transcript (S. cerevisiae, red, T. thermophilus, yellow) are structurally conserved (observed also in transcribing complex PDB:ID 1Y1W). FL1 residues are not present in bacterial polymerases, instead a flexible β′ rudder (residues 583-603, yellow) (47)*

reaches towards the mid-line forming a dome-like structure that spatially organize the junction of the two strands. ***(d)*** *Structure overlay between σ^A factor region 2 (yellow) and open promoter complex (purple) from PDB:ID 4G7O and TFIIB bound to Pol II (PDB:ID 3K7A) (blue). Residues from σ^A region-2 and B-linker differ significantly. While σ^A can recognize specific bases at i-11 and i-7 (DNA scaffold in purple), the B-linker reaches within 6Å to interact with template and non- template strands at the opening end of the transcription bubble and could play a role in scaffold loading.*

3.3.4 Strand separation at the downstream transcriptional fork.

Strand separation in our structure appears to take place at i+5 where base-pair distance begins to increase progressively. However, it was not possible to define its exact location since our artificial bubble enforced non-complementarity precisely at i+5 (see below). However, our structure gives possible insight into the mechanism by which Pol II residues promote and sustain DNA strand separation. The template strand interacts with switch 1 and switch 2 residues driving it towards the active site as originally described (*38*), while the non-template strand interacts with two groups of Pol II residues. The first involves a positively charged cleft formed by Rpb1 residues Arg^{1386}-His^{1387}-Arg^{1391} (switch-1), and Lys^{1102}-Lys^{1109}-Asn^{1110}, (located in a structurally conserved U-loop) in the vicinity of non-template strand i+5 and i+6 (Figures 16A and 17A, (*45*, *60*)). The second involves interactions with Rpb2 fork loop 2 (FL2) residues 501-510 (Figure 16A) providing packing contacts with the non-template strand i+3 to i+1, and Arg^{508} reaching within H-bond distance of the non-template phosphate chain (Figures 16B and 17B).

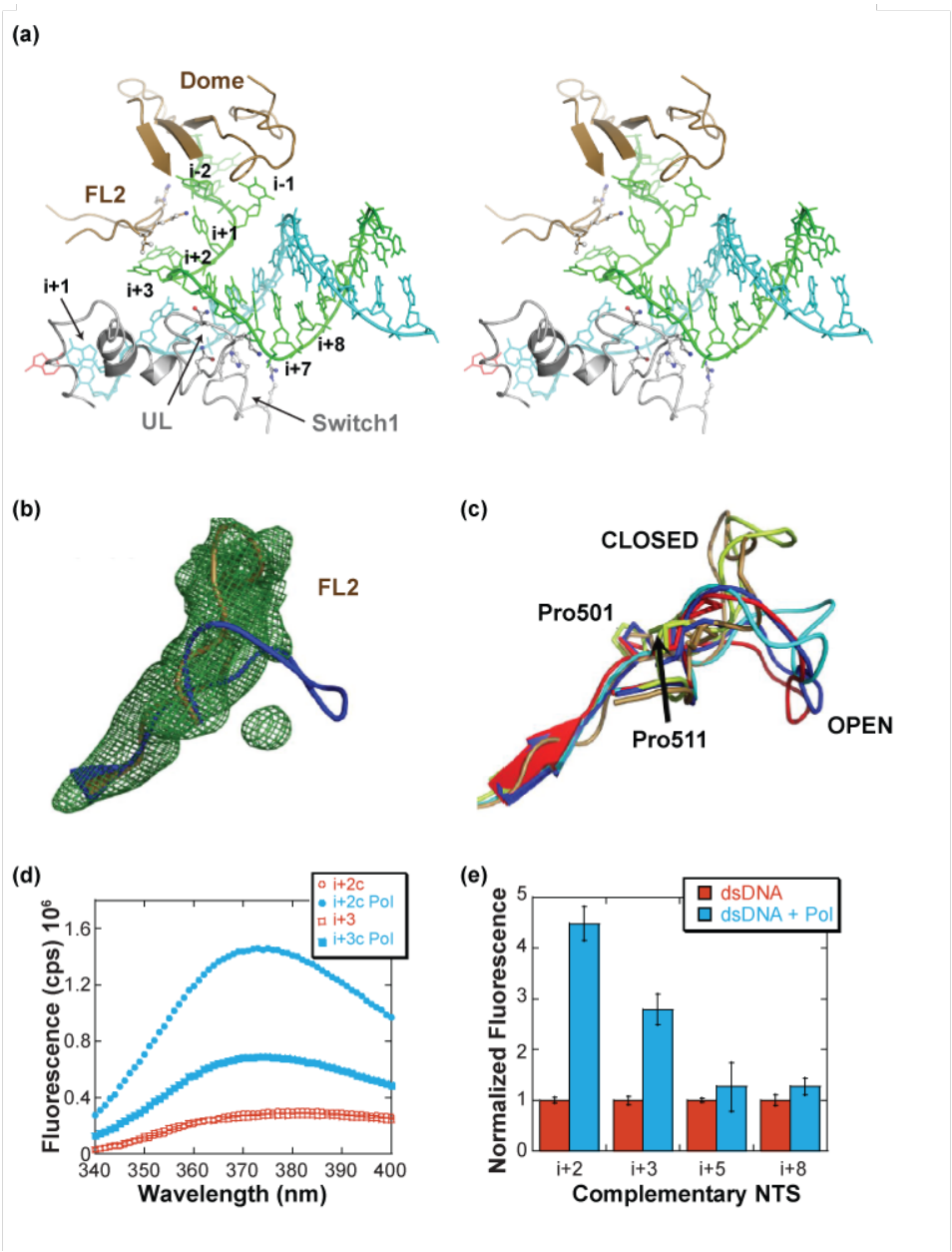

Figure 16. Architecture of the Downstream Fork Junction

(a) *Stereo-view of the architecture of the down- stream fork. Relevant interactions include U-loop (UL) and switch 1 residues with non-template strand i+5 and i+6, FL2 residues with non-tem- plate strand i+2 to i+3, and Rpb2 residues 221– 282 (forming a 5 strand b sheet, Dome) with non- template strand i+1 to i 2.* ***(b)*** *Unbiased Fobs - Fcalc electron density map contoured at 3 s (FL2 residues 498–512 were not included in map calculation).* ***(c)*** *Structural overlay of published FL2 conforma- tions during different stages of transcription. D4/7- TC (sand), PDB: 1Y1W (blue, elongation complex), PDB: 3HOW (cyan, backtracked complex), and PDB: 3PO2 (red, backtracked*

complex). ***(d)*** *Representative spectra of 2AP probes at i+2 or i+3, bound to complementary non-template strand in the absence (open red circles and squares, respectively) and presence (blue circles and squares, respectively) of Pol II. The excitation wavelength was 315 nm and the fluorescence emission (shown in counts per second ×10^6) was collected from 340 to 400 nm.* ***(e)*** *Normalized fluorescence values for polymer- ase bound to ssDNA (primer-template) where 2AP is at the i+2, i+3, i+5, or i+8 position (red) and polymerase bound to dsDNA (primer-template annealed to fully complementary non-template strand) (blue). Error bars are SD of the mean (n = 3).*

Structural overlay of FL2 residues from published crystal structures suggests they could be grouped in two major states (Figure 16C). The first one, an "open" state represented by PDB:IDs 3PO2, 1Y1W and 3HOW, where FL2 residues interact with NTS i+2 or i+3, respectively, allowing access to a non-specific nucleotide-binding pocket (*48*). The second, a closed state where FL2 residues appear to rotate about Pro^{501} and Pro^{510} (our structures, PDB:IDs 3FKI and 3K7A) blocking access to the pocket. Positioning of FL2 residues in the latter conformation appear to guide the non-template strand under a 4-strand β-sheet dome (Figure 16A) towards its junction above the arch. Other interactions include contacts with Rpb2 β-strand 245-255 residues (Figure 17B) and a patch of positive charges from the Rpb2 protrusion helix (Figure 10C).

3.3.5 Fluorescence and structural experiments suggest that the downstream fork is dynamic.

Given that the precise location of strand separation was not possible to be determined from the structure –and in light of previous structural studies showing a closed bubble at i+2 and i+3, respectively (*45, 60*), as well as fluorescence studies where the use of a 35 nucleotide scaffold with a partial (17 nucleotides) non-template strand showed complementarity at i+2 (*120*)– we

wished to ascertain whether it was possible to detect an open bubble in the presence of a 45 nucleotide scaffold bearing a fully complementary non-template strand (Figure 17C). 2-aminopurine (2AP) is a fluorescent nucleotide analog that is significantly quenched upon base pairing to either T or C, as well as by stacking interactions with adjacent nucleotides (*122*, *123*). Therefore, we placed 2-AP in the template strand at positions i+2, i+3, i+5 (at the boundary of strand separation) and downstream at i+8, where the two DNA strands were likely to be paired (Figure 17C). As expected when the primer-template is annealed to a complementary non-template strand, full quenching is observed consistent with stable base pairing of 2AP (Figure 16D,E). When polymerase was added to the scaffold with a fully complementary non-template strand, we see a 4.5-fold increase in fluorescence for the i+2 substrate and a 2.8-fold increase in fluorescence for the i+3 (Figure 16D,E). This increase in fluorescence indicates a disruption in stable 2AP base pairing at both locations and could suggest that Pol II can unwind the non-template strand at i+3. There is 1.2-fold fluorescence enhancement for i+5 and none for i+8 suggesting 2AP is more stably base paired at those positions (Figure 16E). Moreover, crystal structures of Pol II bound to scaffolds 2 and 3 bearing non-complementary transcription bubbles with strand separation at i+2 and i+3, respectively, both showed base-pairing at i+5, recapitulating the results of 2-AP experiments (Figure 17D,E).

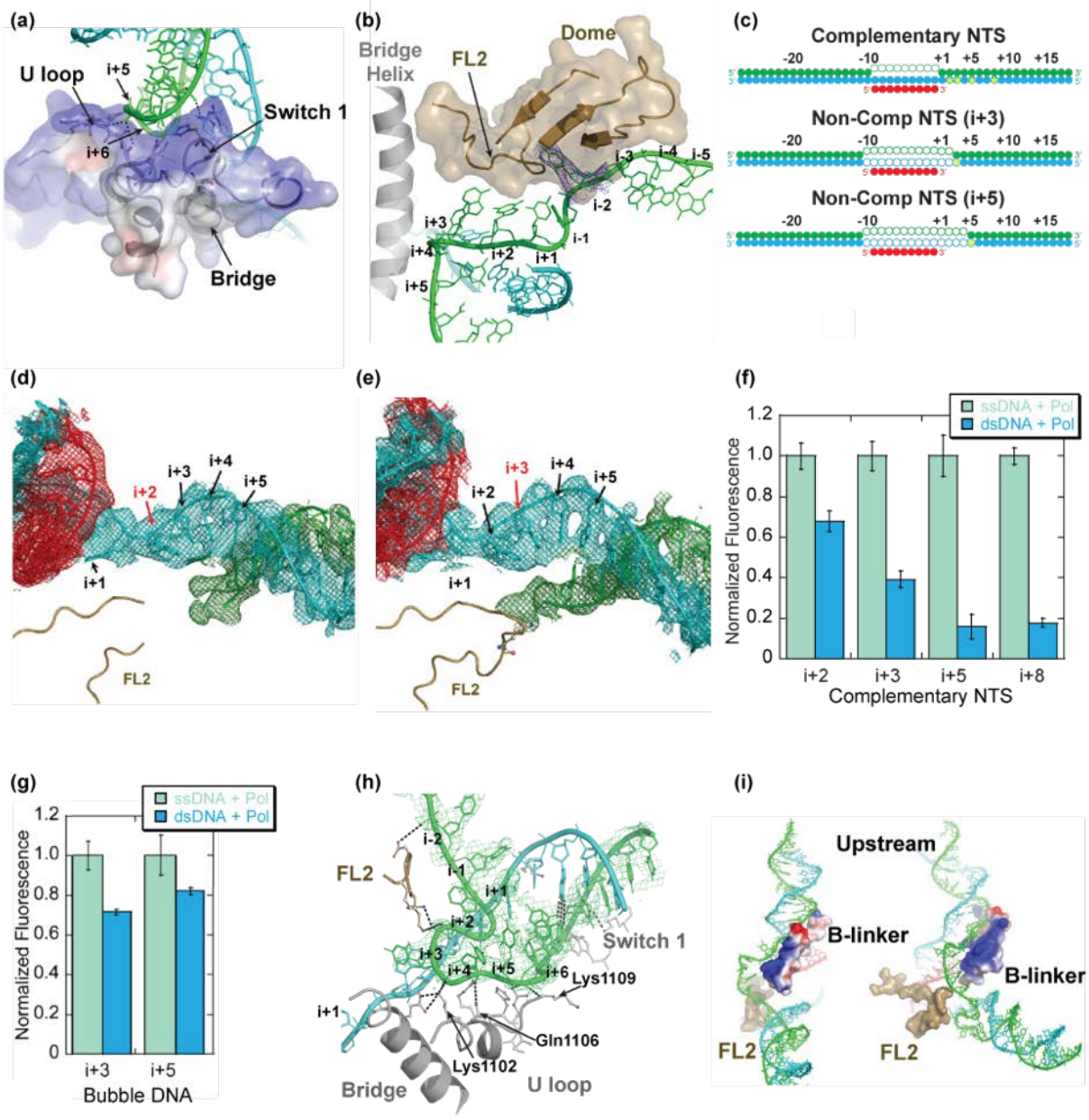

Figure 17. Investigating the dynamics of the downstream fork

(a) *Solvent accessible surface electrostatic representation calculated using APBS (112) in PyMOL to illustrate how non-template strand elements (green) trail a crevice of positive charges (blue) at the opening end of the transcription bubble. Interactions between Pol II elements (U-loop and Switch 1) and the non-template strand phosphate backbone suggest that the region from i+3 to i+6 has the potential to be non-duplexed.* ***(b)*** *FL2 interactions with the non-template strand phosphate backbone lift the strand towards a short tunnel-like structure whose "dome" is a four-strand β-sheet from Rpb2. Electron density for non-template strand position i-2 is shown in*

blue. ***(c)*** *Cartoon schematics of DNA represents the final annealed product used for each 2AP experiment, where yellow stars highlight the position of the 2AP probe, the template strand is in cyan, non-template strand in green, and RNA in red.* ***d,e)*** *Electron density maps countered at 1σ of Pol II transcribing complexes composed of scaffold 2 + Tfg2 (**d**) and scaffold 3 alone (**e**) bearing non-complementary nucleotides at position i+2 or i+3 respectively. FL2 is disordered in i+2 but ordered in i+3 and is primed to engage with the non-template strand at the opening end of the transcription bubble. FL2 residue Arg504 reaches in close proximity to the nucleotide at i+3 position.* ***(f)*** *Normalized fluorescence values for polymerase bound to ssDNA (template/RNA hybrid) where 2AP is at the i+2, i+3, i+5 or i+8 position before addition of the non-template (green) and after addition of non-template strand (blue). Error bars are standard deviation of the mean (n=3).* ***(g)*** *Comparison of 2AP fluorescence against non-complementary (bubble) DNA shows that when non-complementarity is extended to the i+3 or i+5 position, fluorescence measurements increase significantly compared to their complementary counterparts (panel F). Error bars are standard deviation of the mean (n=3).* ***(h)*** *Refined 2Fobs-Fcalc electron density map (green, contoured at 0.8σ) calculated around the opening end of the non-template strand in the Pol II-TC. Evidence of a scrunched state –where template and non-template strand bases are compacted in space relative to relaxed conformations – has been proposed during the early stages of transcription initiation* (*98*, *99*)*. Similarly, other transcriptional events such as backtracking or interactions with elongation or termination factors might alter the number of bases (and location) in the mature transcription bubble.* ***i)*** *Cartoon and surface representation of a front view (left) and front view rotated 60° counterclockwise (right) of the overlay between PDB:ID 3K7A* (*124*) *and Δ4/7-TC. The B-linker region of the general transcription factor TFIIB reaches within 6 Å at the downstream fork junction and could interact with the non-template strand and assist transcription bubble loading during initiation. The electrostatic surface representation calculated using APBS in PyMOL illustrates that one face of the B-linker is positively charged and could possibly allow interactions with non-template strand phosphate chain. FL2 residues could potentially assist during bubble loading via interactions with the non-template strand.*

To observe a full non-template strand it was necessary to utilize a bubble with non-complementary DNA to i+5, since Pol II transcribing complexes bearing non-complementary bubbles at positions i+2 and i+3 showed minimal non-template strand density (scaffolds 2 & 3,

Figure 8A). Thus, we analyzed whether forced non-complementarity at i+5 in the transcribing complex contributed to increased mobility of the non-template strand and its subsequent capture in our scaffold 1 complex. Consistent with this possibility, 2AP experiments of non-complementary bases at positions i+3 and i+5 showed an increased fluorescence signal when compared to bases that are complementary at these positions (Figure 17F,G). This might indicate increased motion of the non-template strand bearing non-complementary bases. It is possible that by forcing non-complementarity at i+5 in our structure, we allowed a kinked conformation of the non-template strand in Pol II-TC structure (Figure 17H). It is also possible, while highly speculative, that the presence of Tfg2 (without Tfg1) in the cleft induced such conformation.

3.3.6 Interactions with the NTS are associated with the off (non-catalytic) state of the trigger loop

A highly conserved loop comprising Rpb1 residues 1078-1097, the trigger loop (TL), has been shown to play a fundamental role in nucleotide selection and catalysis, while also proposed to govern translocation (*49*, *53*, *56*, *58*, *118*, *125*, *126*). Evidence from single molecule studies on Pol II indicates that mutation of TL residues alters Pol II translocation properties, consistent with these models (*126*). Structurally, the TL locates between Rpb1 helices 1064-1078 (TLα_1) and 1097-1106 (TLα_2), which in turn are part of a universally conserved five-helix bundle, (heretofore known as TL bundle (TLB)), that includes Rpb1 helices 826-846 (bridge helix), 1340-1357 (TLα_4) and 1365-1379 (TLα_5) supported by packing of hydrophobic residues at the bundle core (Figure 19A).

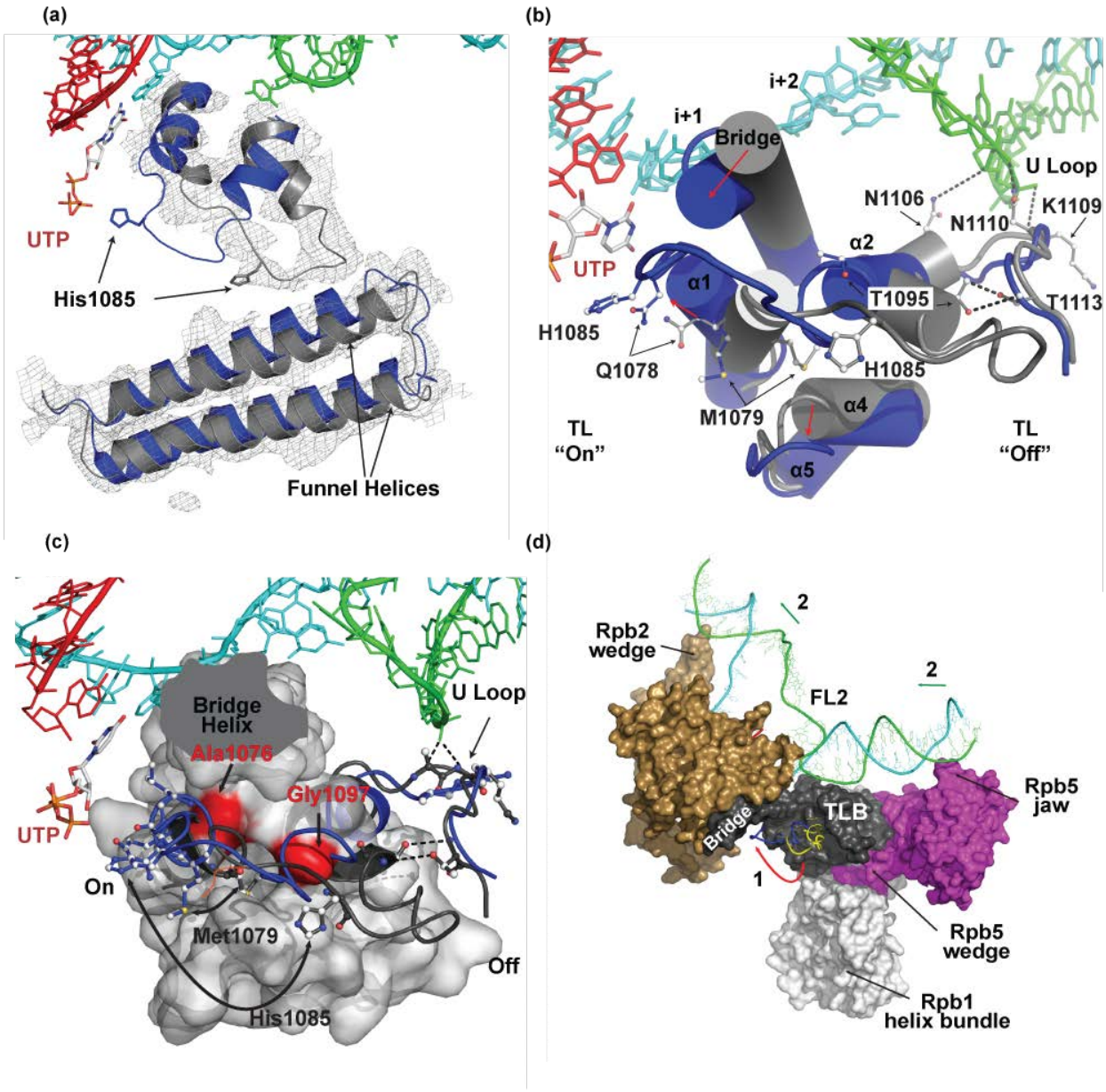

Figure 18. Trigger loop (TL) and nucleic acid scaffold interactions during translocation

On-, off- state residues will be indicated in blue and grey, respectively. A modeled UTP (PDB:ID 2NVZ) is indicated in light grey/orange. ***(a)*** *Conformational changes observed between off and on (structural overlay with PDB:ID 2NVZ) states of the TL and Rpb1 funnel helices. A* $2F_{obs}$-F_{calc} *map rendered at 1.0* σ *is contoured around TL and funnel residues.* ***(b)*** *Conformational changes observed between TL off and on states. Red arrows indicate motion. During off/on state conformational changes, most TL stabilizing interactions are disrupted, including: 1) Release of* Met^{1079} *from its hydrophobic pocket. 2) Disruption of α2-Bridge Helix H-bonds, resulting in bridge helix*

displacement. 3) Disruption of Thr^{1095}-Thr^{1113} H-bonds allowing counterclockwise TL motion. 4) Disruption of non-template strand – U-loop bonds, possibly leading to non-template strand release and translocation. ***(c)*** *Mutations of residues that disrupt Met^{1079} hydrophobic pocket result in gain of function phenotypes* (*56*). *TLB residues are represented as a solid silver surface. Motion of Met^{1079} might occur through a defined pathway on the protein surface (orange trace). Mutations of Ala^{1076} and Gly^{1097} (red surface) for bulkier residues, can potentially disrupt the vestibule of the hydrophobic pocket* (*58*). ***(d)*** *Possible coupling of the global translocation of the scaffold to local motion of the TL. Pol II regions in contact with upstream and downstream duplexes, Rpb2 (sand) and Rpb5 (magenta), respectively, are coupled through TLB residues (dark gray). TL off and on conformations are illustrated in yellow and blue, respectively (see also Figure 19).*

The Pol II TL is intrinsically mobile and has only been detected in X-ray structures when bound to small molecules (such as a matched NTP or α-amanitin) or protein-ligands (TFIIS) that each stabilize a particular conformation (*49*, *57*, *59*, *61*). Observed conformations define a closed "on" state where TL residues interact with a matched nucleotide in the addition or "A" site (effectively isolating a reaction chamber, and blocking access to additional substrates (*58*, *125*)); or an open "off" state where TL residues move away from the A site. During advanced stages of refinement, electron density for the full TL backbone was clearly discernible (Figures 18A and 19B) and partial residue placement was feasible with help of map sharpening in *Coot*. TL backbones were found in novel off states that differ from previously reported conformations (*49*, *61*, *62*) such as the Pol II-TFIIS complex, where direct contacts with TFIIS displaced and stabilized TL residues (Figure 19C). Interestingly, our structures show two distinct off state conformations: The first one present in the Δ4/7-TC closely resembles the TL loop in the "on" conformation (PDB:IDs 2E2H and 2NVZ, (*49*)) but rotated counter-clockwise approximately 60° away from the addition site (Figure 18A).

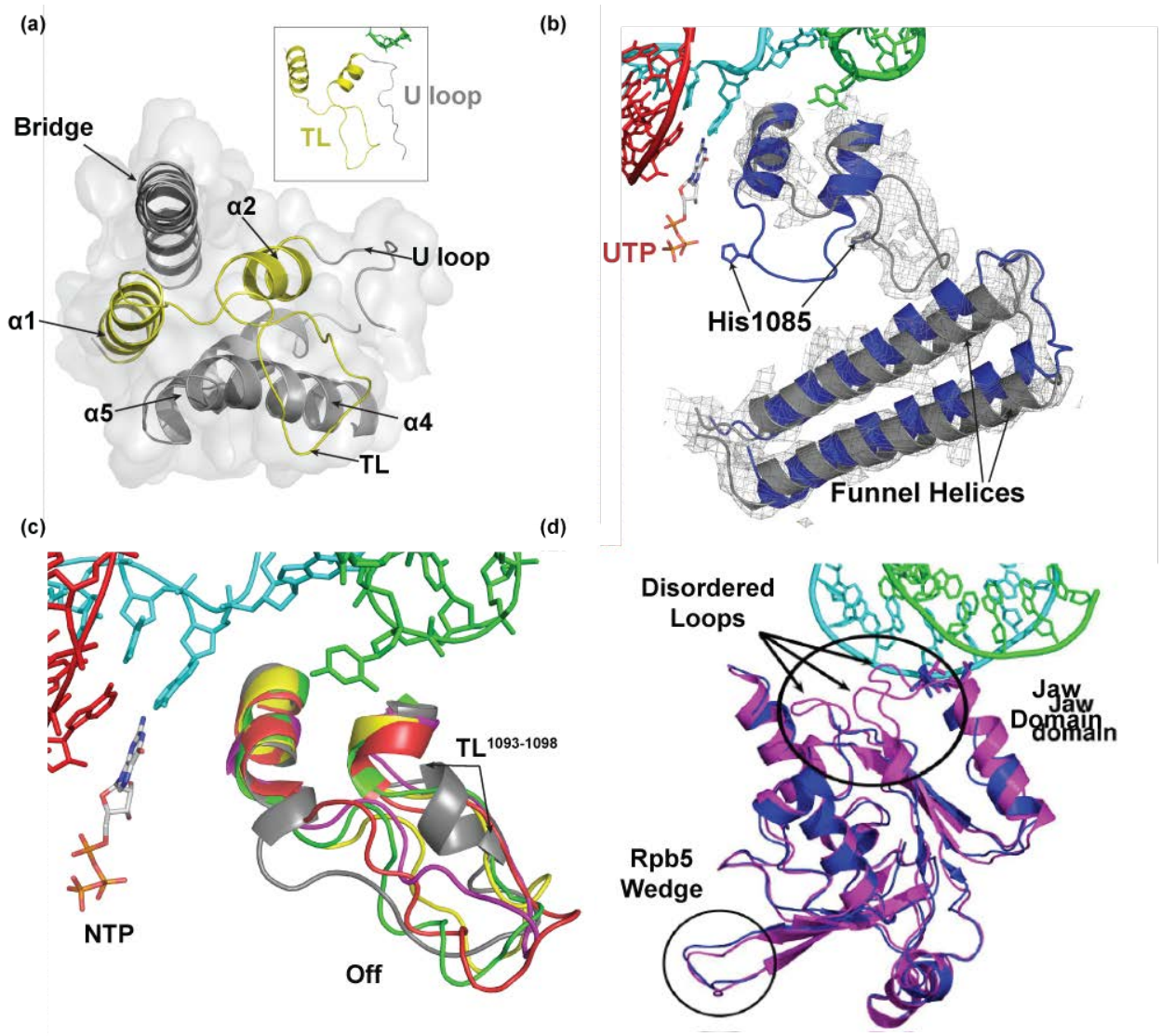

Figure 19. Conformational changes associated with DNA translocation in trigger loop and Rpb5 residues

(a) *Cartoon and surface representation of the Rpb1 5-helical bundle comprising the trigger loop-bundle (TLB). α1, residues 1063-1077; α2, residues 1098-1106; Bridge helix, residues 810-845; α4, residue 1341-1357; and α5, residue 1365-1377. TL, TL helices α1-α2 (hinges), and the U-loop are illustrated in yellow (Inset, also shows non-template strand i+5 and i+6). The core of the bundle is held together by hydrophobic interactions.* ***(b)*** *Conformational changes observed between "off" and "on" (structural overlay with PDB:ID 2NVZ, blue) states of the TL and Rpb1 funnel helices. All "on" state residues are indicated in blue, "off" state residues are indicated in grey for Pol II-TC. A $2F_{obs}$-F_{calc} map rendered at 1.0 σ is contoured around TL and funnel residues.* ***(c)*** *Conformational changes observed between various "off" states of the TL: Δ4,7-TC, silver; Pol II-TC, yellow;*

PDB:ID 1Y1V, red; PDB:ID 3CQZ, magenta; PDB:ID 2NVX, green. Differences in secondary structure for TL residues 1093-1098 in Pol II transcribing structures, suggests that we were able to capture "off" state intermediates. ***(d)*** *Cartoon representation of Rpb5 residues during "off" (illustrated in magenta) and "on" conformations, (illustrated in blue). During "on" state three loops of the jaw domain become disordered, suggesting that this region "moves" during this stage facilitating translocation of the nucleic acid scaffold. Additional conformational changes are observed in Rpb5 wedge residues. These changes could be related to motion in TLB residues.*

The second one, observed in Pol II-TC resembles a hairpin loop and its shape is similar to previously reported off state structures (Figure 19A-C, (*45*, *49*)). Both conformations induce considerable changes in funnel and neighboring residues (Figures 18A and 19B). Albeit these differences in the positions of the TL's off state, a common set of interactions can be observed among these structures and all structures crystallized with a DNA scaffold (exclusive of the on-state structures). These include: 1) interactions between TL hinge regions (TLα1 and TLα2 helices) with neighboring Rpb1 residues, and 2) the burying of Met^{1079} inside a small hydrophobic pocket at the core of the TLB (Figure 18B).

Structural overlay between off and on states suggest substantial rearrangements in TLB helices and TL residues during the nucleotide addition cycle (Figure 18B). Transition to the on state involves TL residues swinging towards the addition site, where hinge contacts are disrupted and Met^{1079} moves out of its pocket at the core of the TLB (Figure 18C). Genetic interactions and gene expression profiling of substitutions of Rpb1 Ala^{1076}, Gly^{1097} and Leu^{1101} with residues that disrupt the Rpb1 Met^{1079} hydrophobic pocket (*127*), support a model where the integrity of the pocket is critical for stabilization of the off state in eukaryotic (*56*, *57*) and archaeal (*55*) polymerases. Such substitutions are highly related to those that hamper off state conformations by destabilization of the C-terminal TL hinge region; therefore, these substitutions are similarly

predicted to alter translocation rate and catalysis (*49*, *56-58*, *128*) suggesting that the TL "off" state is specifically required for proper transcription.

3.3.7 Structurally linked DNA-interacting domains: Upstream (Rpb2), downstream (Rpb5) and TLB residues could possibly coordinate NAS translocation

Three regions of Pol II furnish residues that can potentially form H-bond contacts with the nucleic acid scaffold (Figure 20A). These include (from upstream to downstream): 1) wedge, wall and FL1 (Rpb2) with the upstream duplex, template strand and the non-template strand (respectively); 2) TLB, rudder, switch (1 and 2) and clamp residues (Rpb1), with DNA-RNA hybrid, template strand, downstream fork and downstream duplex (respectively); and 3) jaw residues (Rpb5) with the downstream duplex (Figures 18D and 20A,B). Remarkably, these regions have substantial interactions among them: TLB helices ($TL\alpha_4$ and $TL\alpha_5$) and Rpb1 residues from a seven-helix bundle ($Rpb1^{846\text{-}1064}$) form a large pocket that buries a two-strand hairpin from $Rpb5^{193\text{-}214}$, which is in the immediate neighborhood of the jaw motif. Similarly, TLB contacts (via the bridge helix) with Rpb2 are extensive and involve "wall" residues (with the phosphate chain of the template strand) in the immediate neighborhood of Rpb2's wedge (Figures 11B and 18D). Since these regions are coupled extensively, it is possible that they could play an important role during translocation.

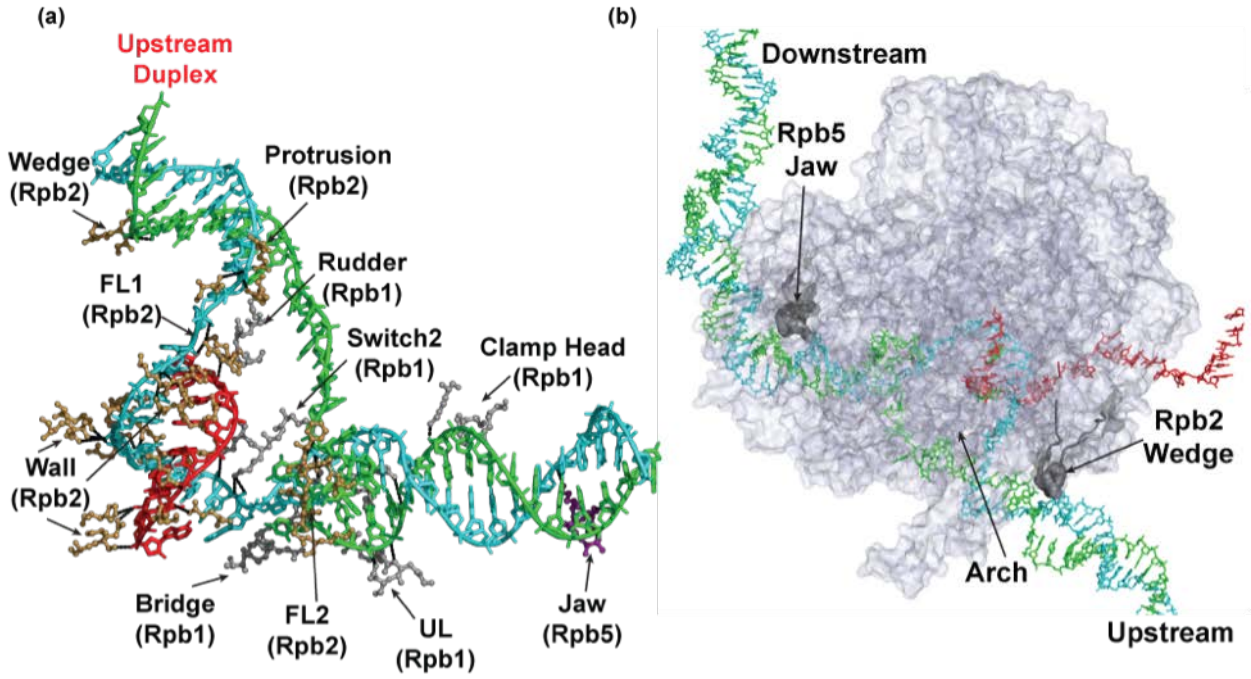

Figure 20. Network of Pol II - NAS interactions

(a) *Illustration of residues that potentially contribute to hydrogen bond stabilization of the nucleic acid scaffold. Rpb1, Rpb2 and Rpb5 residues are illustrated in silver, copper and magenta respectively. Over 40 potential H-bonds are disrupted during each translocation cycle (H-bond distances of 3.5 Å or less).* ***(b)*** *Surface representation of a transcribing Pol II complex. Cyan regions highlight interactions between Pol II and the nucleic acid scaffold, which may comprise the structural framework for high fidelity DNA tracking.*

3.4 DISCUSSION

3.4.1 Molecular basis for high-fidelity DNA-tracking

Real time microscopy experiments demonstrated the ability of RNA polymerases to rotate DNA by tracking with high fidelity its right hand helix (*129*). Our structures suggest that it is possible that engagement by wedge (upstream duplex), arch (closing end of the bubble) and jaw and clamp (downstream duplex) residues (Figures 10, 13 and 20F) could comprise the structural framework that explains such tracking mechanism. The interactions that Rpb1 (head clamp) and Rpb5 (jaw) residues make with the downstream duplex were described in several published crystal structures (*43*, *49*). However, since contacts with the upstream duplex and arch were not previously observed, the correlation between DNA tracking, and its structural underpinnings could not be established. Moreover, given that the crystal structures of Pol II transcribing complex and mtRNAP show interactions with upstream and downstream duplexes, it is possible that tracking mechanisms are conserved in transcription (Figures 11C and 12). Furthermore, overlay of the archaeal RNAP clamp domain in complex with the heterodimer interface of the Spt4/5 complex (PDB:ID 3QQC, (*130*)) with the Rpb1 clamp domain of Δ4/7-TC shows that the upstream duplex is situated between the wedge domain and position of the Spt5 NusG domain (Figure 21A). While speculative, the observed location of the duplex in our structure is consistent with previous biochemical data and Pol II – Spt4/5 elongation models (*130*-*132*).

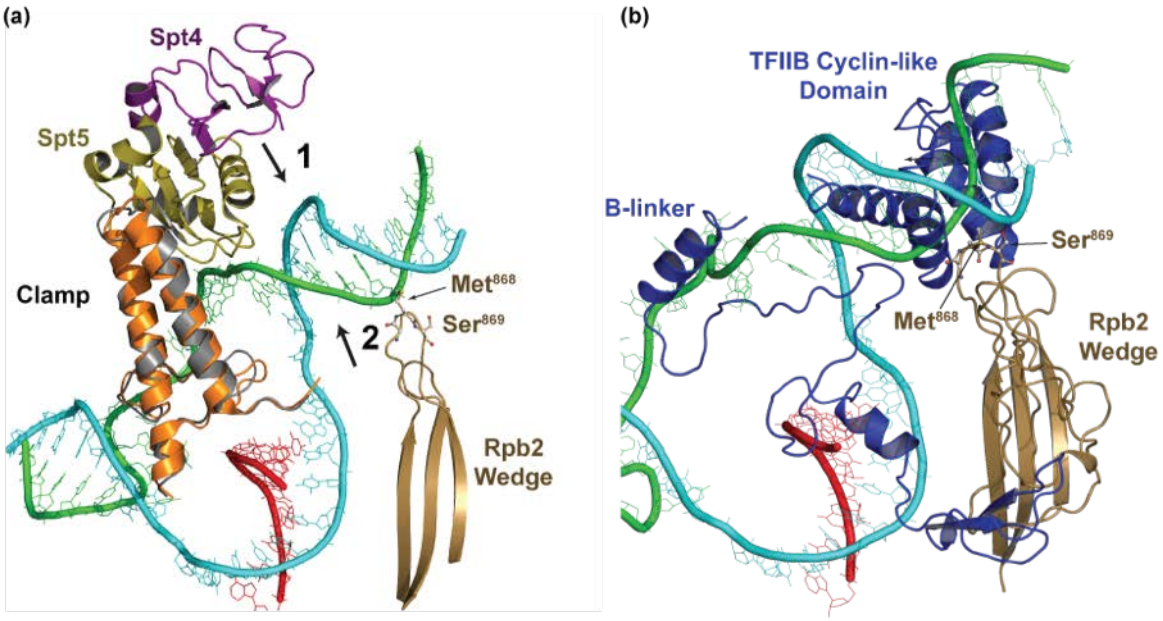

Figure 21. Modeling NAS interactions with initiation and elongation factors

(a) *Overlay of Δ4,7-TC with the archaeal P. furiosus clamp-Spt4/5 complex, PDB:ID 3QQC (Martinez-Rucobo et al., 2011) about the clamp coiled coil domains of Pol II Rpb1 (grey) and RNAP A' (orange), suggests that the upstream duplex is positioned between the NusG domain of Spt5 (olive) (arrow 1) and Rpb2 wedge domain (arrow 2). While speculative, our model suggests that Spt4/5 locks the upstream duplex in place and may play a role in maintaining bubble integrity during elongation. This is in agreement with previously published models* (*130*, *131*)*.* ***(b)*** *Structure overlay with initially transcribing Pol II-TFIIB structure, PDB ID: 4BBS* (*81*) *illustrate possible interactions between wedge residues (copper) and Cyclin-like domain residues (blue). Comparisons between a model of the closed promoter complex* (*61*, *62*) *and the structure of the transcribing complex suggest that transition from closed to open promoter entails a clockwise rotation (approximately of 90º degrees) of the upstream end of the DNA. Wedge residues Met*^{868} *and Ser*^{869} *interact with TFIIB residues 148-153 and 178-183 from the cyclin-like domain, such interactions could participate in Pol II duplex loading. It is possible that this rotational motion induces steric clashes that ultimately result in ejection of TFIIB and TBP allowing the axle to engage the minor groove. The close proximity of these two regions could allow loading of the duplex on the wedge during the initiation to elongation transition*.

In collaboration with the lab of Dr. Craig Kaplan, we assessed multiple *rpb2* wedge alleles to ascertain the role of these elements *in vivo*. The strongest growth effects observed *in vivo* required removal of the loop (Figure 12D), which may be considered an extensive perturbation to the Pol II structure. However, our interpretation of the direct or indirect functions of this loop is based on what is known about other Pol II mutants. We found that wedge alleles were MPA sensitive and shifted start sites downstream (Figures 11H and 12D). This profile is relatively rare for Pol II alleles or known general transcription factor alleles (reviewed in (*118*)), especially for mutants unrelated to the active center (Rpb1 N488D within the Pol II active site has this phenotypic profile (*128*)). Recent crystal structures of Pol II-TFIIB structures show that the wedge domain interacts with the TFIIB core N-terminal cyclin fold (*81*). Since TFIIB is necessary for start site selection, it seems plausible that changes in Pol II-TFIIB interactions could affect this process. Deletion of the wedge loop confers a strong defect in occupancy of Pol II at a reporter gene, consistent with an initiation defect. However, the wedge mutants do not phenocopy TFIIB (*sua7*) alleles in relation to MPA-sensitivity or *IMD2* expression (Figure 12D), suggesting that some functions may be independent of TFIIB or relate to a novel initiation defect not observed in particular *sua7* alleles. The wedge domain may possibly have a role during transition from initiation to early elongation by providing interactions with the upstream duplex that could assist TFIIB ejection (Figure 21B).

3.4.2 Downstream fork flexibility

The downstream fork is a dynamic region and different conformations of the non-template strand could be allowed during different stages of transcription. Structural overlay of the non-template strand from our structures and published structures crystallized in the presence of a nucleic acid

scaffold show small positional differences from i+5 to i+7, which are anchored by U-loop and switch 2 residues (Figure 17A). However, positions i+2 to i+4 differ in all crystal structures and can be found paired or un-paired to the template strand. These observations suggest that there may be a range within which the leading edge of the downstream fork can move; hence, fluidity within this region could be required to promote bubble opening and maintenance. Moreover, non-template strand flexibility correlates with FL2 flexibility (Figure 16C), which adopts multiple conformations (*48*, *49*, *80*, *103*) to allow interactions with the non-template strand during the various transcriptional stages. The size of the bubble itself, which appears to be dynamic, might not be so critical due to the fact that Pol II can accommodate different lengths and possibly different locations of the transcription bubble inside the cleft (*82*).

The crystal structure of the ternary complex between *T. thermophilus* polymerase (RNAP), σ^A and an open promoter complex (*133*) shows that FL2 residues fold to form part of a 2′-deoxy-GTP "selectivity-pocket" for non-template strand i+2 during initiation. Also, the crystal structure of the RNAP ternary complex revealed that promoter recognition, melting and bubble loading are carried out by σ^A region-2 residues (Figure 15C). Thus architecture of the downstream fork in RNAP during initiation is determined by σ^A as well as FL2 residues.

The recent crystal structure of *E.coli* RNAP holoenzyme in complex with a 15-nucleotide bubble at 6 Å resolution revealed for the first time the architecture of the RNAP transcription bubble stabilized by σ elements (*134*). However, overlay of RNAP bubble structure with our structures show different trajectories of the non-template strand and also different position of the upstream duplex. Such differences could be explained structurally since, on the one hand FL1 residues (not conserved in RNAP) would clash with the non-template strand of RNAP; and on the other, the presence of σ, which itself interacts with the non-template strand, in the RNAP

cleft shifts dramatically the position of the upstream duplex.

Interestingly, overlay of the co-crystal structures of Pol II in complex with TFIIB (*74*, *79*, *80*) with the our structure shows that TFIIB linker domain reaches within 6.5 Å of the fork junction and could assist bubble loading (*80*, *81*) and prevent re-annealing of downstream fork during initiation (Figure 17I). Thus, σ^A and TFIIB could sustain an open fork (during initiation) using different mechanisms.

3.4.3 TL allosteric effects and DNA translocation

It has been proposed by a number of groups that TL movement contributes or controls translocation (*53*, *56*, *59*, *125*, *126*). Our structures show Pol II in a post-translocated state, with a TL in the off state due to interactions with neighboring Rpb1 residues (Figures 18A,B and 19B). It is possible that a matched NTP at the i+1 position might disrupt these interactions leading to an on state. The on-state structure (*49*) shows Gln^{1078} as one of the key residues stabilizing a matched nucleotide through formation of H-bonds with the ribose. Structural overlay of TL residues between off and on conformations shows that Gln^{1078} moves approximately 3 Å to form H-bonds with the matched nucleotide (Figure 18B, see also (*60*). The position of this residue might be critical since on the one hand, it could constitute part of the nucleotide selection mechanism (*55*, *135*). On the other hand, displacement of Gln^{1078} could trigger extraction of the neighboring Met^{1079} from its hydrophobic pocket, initiating a cascade of events that would lead into the full on state. Along with our observations, genetic evidence has shown that mutations of Gln^{1078} have a comparable effect on Pol II elongation activity as mutations on the catalytic His^{1085} (*56*). Importantly, Gln^{1078} and His^{1085} are genetically distinguishable suggesting a multistep process in TL function, consistent with initial substrate-Gln^{1078} interactions and

subsequent TL movement or folding. Moreover, Pol II activity is also exquisitely sensitive to substitutions around the Met^{1079} hydrophobic pocket (Figure 18C) (*56, 58*). These substitutions invariably lead to genetic phenotypes, genetic interaction and gene expression profiling phenotypes consistent with increased Pol II activity, most likely due to destabilization of the TL off state (*127*).

Critical questions for the Pol II mechanism are how translocation occurs and what are the molecular determinants of its linkage to the nucleotide addition cycle? Our structure reveals that a second pivotal role played by TLB residues could possibly include stabilization of the non-template strand. Crystal structures of Pol II bound to partial or full nucleic acid scaffolds (*43*, *45*, *49*, *60*) show that TLα_2 and UL residues Lys^{1102}, Asn^{1106}, Lys^{1109} (amide backbone) and Asn^{1110} locate within H-bond distance to the phosphate chain of non-template strand i+5 and i+6 (Figure 17A). Overlay of the two states shows that the hinged motion of the TL during matched nucleotide binding results in conformational changes leading to increased U-loop – non-template strand distance and hence disruption of potential H-bond contacts (Figure 18B,C). Although speculative this could suggest that TL “off/on” transitions are allosterically coupled to “latch and release” (respectively) events of the non-template strand phosphate chain by U-loop residues. Furthermore, comparisons between on and off state conformations show that on-state Rpb5 jaw residues (interacting with the downstream duplex) become disordered, hence decreasing the number of effective contacts with the duplex (Figure 19D). Since TLB, Rpb5 (jaw residues) and Rpb2 (wall, wedge) regions could be coupled, through extensive observed interactions, it is possible that TL conformational changes are transferred allosterically to downstream and upstream duplexes to assist global translocation (Figures 18D and 20).

Our refined structure sheds light into four fundamental mechanistic aspects of

transcription: 1) In addition to hybrid interactions (*38*), Pol II has four major contact points with a nucleic acid scaffold: wedge, arch, clamp and jaw domains (Figure 10A,B). Engagement of the minor groove by Rpb2 wedge (upstream), arch (closing end of the bubble) jaw and clamp (downstream) residues could comprise the structural framework that explains the mechanism for high fidelity DNA "tracking" observed using real-time optical microscopy (*129*). 2) Pol II residues define the architecture of the transcriptional fork. Arch residues coordinate annealing of template and non-template strands at the upstream fork; the downstream fork is a highly dynamic area where FL2 residues accommodate different conformations of the non-template strand. 3) The structure also suggests that Gln^{1078} is positioned to couple extraction of Met^{1079} out of the hydrophobic pocket to interactions with an incoming matched NTP, initiating a cascade of events leading to a full on state of the TL, followed by nucleotide incorporation and subsequent DNA/RNA translocation. 4) Finally, our structure shows that Pol II regions in contact with the nucleic acid scaffold are connected as rigid bodies (from Rpb1, Rpb2 and Rpb5) and that TL on/off state conformational changes could possibly be tied to global translocation.

3.5 ACKNOWLEDGEMENTS

We would like to acknowledge the user support of Michael Becker and Craig Ogata at GM/CA (Argonne National Laboratory) and to Irimpan Mathews, Clyde Smith and Ana Gonzalez at Stanford Synchrotron Radiation Lightsource (SSRL), for their magnificent support during data collection. COB acknowledges support from NIH T32GM008424. GC acknowledges University of Pittsburgh startup funds and NIH R01 GM112686. CDK acknowledges support from NIH R01 GM097260 and Welch Foundation Grant A-1763. Conclusions are the sole thoughts and opinions of the authors listed.

4.0 STRUCTURAL STUDIES OF MULTI-PROTEIN POL II COMPLEXES: CHALLENGES AND DEVELOPMENTS

While macromolecular complexes are biologically prevalent and important, their structural data is sparse when compared to isolated single proteins (*136*). To perform and regulate critical cellular processes, such as transcription, proteins form interacting heterogeneous assemblies (multi-protein complexes or MPCs). Estimates for the number of interacting protein pairs ranges from ~7,000 in *E. coli* to greater than 45,000 in *S. cerevisiae*, (*137*) whereas in higher order organisms it may exceed half a million (*138*). These interactions form the basis of the quaternary structure of MPCs and represent one of the most complex levels of structural organization in biological molecules. Out of the more than 100,000 entries in the PDB, just over 8% are protein-protein or protein-nucleic acid complexes with a fraction of that representing complexes of three or more proteins (*139*). The structural specificity of protein-protein interactions is of crucial importance to fully understand the underlying mechanisms that govern cellular functions, and in the case of the transcription machinery, provide a blueprint for structure-guided therapeutics (as discussed in Chapter 1). While the advent of single-particle cryo-EM techniques have successfully allowed the study of MPCs from low – medium resolutions (see section **2.1.5**), x-ray crystallography still presents the best opportunity at garnering the atomic-level details along the protein-protein interfaces for therapeutic design. Therefore, examination of the bottlenecks

that prevent MPC crystallization and methods to overcome these challenges will be discussed in the proceeding chapters.

4.1.1 Methods to overcome inherent challenges with MPC crystallization

Co-crystallization of Pol II with the various TFs that regulate its activity has been limited to only a handful of high-resolution structures. The stubbornness of complexes to crystallize may be in part due to conformational disorder, or flexibility often associated with structural rearrangements involved with function and/or complex formation (*140*). This is not unique to Pol II complexes, as conformational flexibility resulting in unstructured regions of 40 amino acids or more in length, is present in at least 50% of eukaryotic proteins (*141*) and significantly correlates with the probability of crystallization failure (*142*). In addition, affinities between the individual components of MPCs are weak, due to the transient nature of their interactions, and often result in complex disassembly during the crystallization process. Such incorrect stoichiometry can interfere with crystallization; evident by early Pol II crystallization experiments where sub-stoichiometric Rpb4/7 led to poor crystal growth (see section **2.1.1**).

Indeed, complex integrity is likely the main reason that MPCs are more difficult to crystallize than individual soluble proteins. A restricted set of chemical conditions must both stabilize the MPC and simultaneously promote the weak interparticle interactions required for the spontaneous nucleation of the MPC into an ordered lattice (Figure 22). While the boundaries of chemical space are reported to be more restricted (*143*) the sensitivity to this space is far higher. Considering that the majority of crystallization screens were developed for the crystallization of single, soluble proteins, it is not surprising that these same screens are inherently less effective for the crystallization of MPCs. Thus, a major effort is underway to

develop methods which improve complex stability, and possibly expand the chemical space for MPC crystallization.

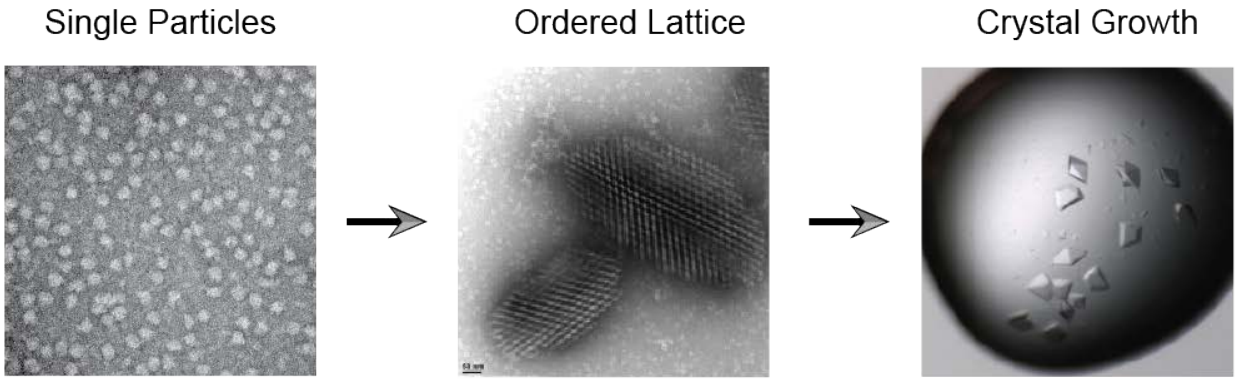

Figure 22. The crystallization process

A flow scheme, representing the crystallization process of Pol II. The spontaneous nucleation of homogeneous Pol II particles (left panel) into an ordered crystal lattice (middle panel) is often the rate-limiting step in crystallization experiments. Conditions which promote growth lie at the interface of nucleation, and must be optimize to produce large, single-diffracting crysals (right panel). Degradation of sample overtime can prevent crystal growth, limiting crystals to nanometer sizes (middle panel).

Advances in molecular biology and biochemical techniques have allowed for the expression, purification and stoichiometric reconsitution of MPCs and pharmacologically relevant protein targets (*144-146*). Similarly, innovations in synchrotron technology, including development of micro-focused X-ray beams, advances in X-ray area detectors, as well as the development of user-friendly crystallography software have significantly shortened crystal-to-structure time frames. Recent efforts towards improving sample "crystallizability" using techniques such as alanine scanning mutagenesis (*147-150*) or the use of chimeric proteins to promote/improve crystal packing (*150*, *151*) have also led to structures of important targets (*148*, *151-155*).

The more common crystallographic practice of crystal microseeding, where crushed

crystals are introduced into new drops using a probe, has also been instrumental in promoting crystal growth (*156-158*). This technique, along with similar adaptations like microseed matrix screening (MMS) or cross-seeding using one component of a MPC (*159*, *160*), has proven useful for the production and optimization of high-quality, diffracting crystals. Furthermore, the emergence of synthetic Fab fragments to stabilize MPCs which have avoided crystallization has shown great promise at expanding the chemical space for crystallographic trials (*161-164*).

4.1.2 TEM as a tool to guide crystallography experiments

Even if crystallization is successful, diffraction from MPCs is problematic. Crystals are typically small, fragile, and difficult to physically manipulate, and present additional challenges that include twinning and diffraction that can be both anisotropic and limited to low resolution. In addition, a major hurdle is the inability to identify conditions that promote crystallization, especially when crystal growth is limited to nano-meter sized dimensions. Emerging techniques utilizing second-order nonlinear optical imaging of chiral crystals (SONICC, (*165*)), UV-microscopy (*166*), and negative-stain TEM (*167*, *168*)) have shown that protein nanocrystals are ubiquitously present in granular aggregates of crystallization drops, which would usually be overlooked by the experimenter. Using TEM we showed that in many occasions, such nanocrystals contained well-ordered lattices (*168*), which may represent ideal material for novel injector-based serial femtosecond crystallography (SFX) experiments at x-ray free electron laser (XFEL) sources or micro-electron diffraction (microED) experiments. The advent of these techniques has significantly altered the field of structural biology, as atomic- resolution structural information can now be obtained from nanocrystals (*169-173*). However, limited access to these instruments and the preparation of crystalline samples still represents a major roadblock for

experimenters.

In the proceeding chapter, we build on our previous work in identifying protein nanocrystals and employ negative stain TEM as a tool to study and optimize crystallization in three ways. First, we use a fragmentation protocol applied to crystals or nanocrystals (observed in UV-positive granular aggregates) to generate crystal fragments. The detection of Bragg spots calculated from TEM images of such fragments correlates positively with successful X-ray diffraction. Second, we demonstrate the benefits of TEM to study the process of crystallization, from sample analysis to estimations of solvent content to crystal optimization (i.e., TEM-guided crystal growth). Third, we establish the usefulness of high quality crystal fragments from crystals or nanocrystals, to grow reproducibly larger size crystals.

5.0 TRANSMISSION ELECTRON MICROSCOPY FOR THE EVALUATION AND OPTIMIZATION OF CRYSTAL GROWTH

5.1 SUMMARY

Crystallization of protein samples remains the most significant challenge in structure determination by X-ray crystallography. Here we demonstrate the effectiveness of transmission electron microscopy (TEM) analysis to aid in the crystallization of challenging protein targets and MPCs, including several Pol II transcribing complexes. We found the presence of well-order lattices with higher order Bragg spots, revealed by Fourier analysis of TEM images, as a good predictor of diffraction-quality crystals. Moreover use of TEM allowed 1) comparison of lattice quality among crystals from different conditions in crystallization screens; 2) detection of crystal pathologies that could contribute to poor X-ray diffraction, including crystal lattice defects, anisotropic diffraction and crystal contamination by heavy protein aggregates and nanocrystal nuclei; 3) qualitative estimation of crystal solvent content to explore the effect of lattice dehydration on diffraction; and 4) selection of high quality crystal fragments for microseeding experiments to generate reproducibly larger size crystals. Applications for X-ray free electron laser (XFEL) and micro electron diffraction (MicroED) experiments are also discussed.

Work presented in this chapter was reproduced with permission from the International Union of Crystallogarphy:

Stevenson HP*, Lin G*, **Barnes CO***, Sutkeviciute I, Krzysiak T, Weiss SC, Reynolds S, Wu Y, Nagarajan V, Makhov AM, Lawrence R, Lamm E, Clark L, Gardella TJ, Hogue BG, Ogata CM, Ahn J, Gronenborn AM, Conway JF, Vilardaga JP, Cohen AE, Calero G. (2016) Applications of TEM for the evaluation and optimization of crystal growth. *Acta Cryst.* **D72(5),** 603-15. doi: 10.1107/S2059798316001546

5.2 EXPERIMENTAL METHODS

5.2.1 Protein purification, crystallography, and UV fluorescence screening

We utilized crystals derived from various protein samples, which were categorized into four different protein classes: 1) Soluble, 2) Membrane, 3) Multi-protein Complex, or 4) Viral. References for protein purification and crystallization conditions are summarized in Table 4, which was conducted in collaboration with multiple laboratories.

Table 4. Protein Purification and Crystallographic Conditions

Protein[a]	Class	Crystallization Condition[b]	Avg. Order Bragg Spot	Max[c] Resolution (Å)
Pol II Δ 4/7 (Cramer *et al.*, 2001)	Soluble	0.4 M Ammonium/Sodium Phosphate pH 6.5, 6-10% PEG 6000, 50 mM Dioxane, 10 mM DTT	3rd	3.2
Wt Pol II (Pullara *et al.*, 2013)	Complex	0.1 M HEPES pH 7.0, 0.3 M Sodium Malonate, 0.2 M Ammonium Acetate, 3-6% PEG6000, 10 mM DTT	3rd	3.5
Pol II – TFIIF – DNA (Pullara *et al.*, 2013)	Complex	0.1 M HEPES pH 7.0, 0.1 M Sodium Malonate, 8-12% PEG4000, 10 mM DTT	3rd 1st	3.8 (Aniso) 5.0
Pol II – Spt4/5 – DNA (Cohen *et al.*, 2014)	Complex	0.1 M MES pH 6.0, 0.2 M Magnesium Chloride, 10-14% PEG2000, 10 mM DTT	3rd 2nd	3.9 (Aniso) 4.5
Pol II – TFIIB – DNA (Cohen *et al.*, 2014)	Complex	0.1 M HEPES pH 7.0, 30-35% Tacismate pH 7.0, 10 mM DTT	N/A	3.4
Pol II – GFP (Cohen *et al.*, 2014)	Complex	0.1 M HEPES pH 7.0, 0.8 M Sodium Citrate, 10 mM DTT	3rd 2nd	3.6 (Aniso) 4.3
Pol II – CD3Δ (Cohen *et al.*, 2014)	Complex	0.1 M MES pH 6.0, 1.4 M Sodium Malonate, 10 mM DTT	3rd 2nd	4.4 (Aniso) 5.7
UNG2-DCAF-DDB1-Vpr (Ahn *et al.*, 2010)	Complex	0.1 TRIS pH 7.5, 0.1 M Sodium Acetate, 15% PEG 4000	3rd	3.7
DCAF-DDB1-Vpr (Ahn *et al.*, 2010)	Complex	2.0 M Sodium Chloride, 12 % PEG6000	3rd	3.6
PTHR (Pullara *et al.*, 2013)	Membrane	0.1M Sodium Phosphate, 12% PEG4000, 5 mM TCEP, pH 6.2	2nd	**N/A**
BRIL-PTHR[d]	Membrane	0.1 M Sodium Citrate, 10% PEG8000, 0.1 M Calcium Chloride, 10 mM DTT, pH 5.8	1st	**N/A**

2,3 – HPCD (Botha *et al.*, 2015)	Soluble	0.1 M MES pH 5.8, 0.1 M Calcium Chloride, 12-14% PEG6000	4th	1.55
Globin-X[e]	Soluble	0.2 M Sodium Malonate pH 7.0, 20% PEG3350	1st	<10
APOBEC Adapted from (Byeon *et al.*, 2013)	Soluble	0.1 M MES pH 6.2, 1.5 M Sodium Citrate	1st	<10
		0.1 M HEPES pH 7.0, 0.2 M Magnesium Acetate, 10% PEG4000	2nd	6.3
Rtf1 (Dermody & Buratowski, 2010)	Soluble	0.1 M MES pH 6.0, 0.1 M Magnesium Chloride, 15% PEG3350	1st	7.3
dGTPase[f]	Globular	1.6 M Potassium/Sodium Phosphate pH 8.5	4th	2.8
Sindbis Virus	Viral		1st	<10

[a]Purification and complex assembly protocols were performed as previously described for all proteins or complexes where a reference is provided.

[b]In all cases, crystallization was achieved using vapour diffusion with drop sizes between 1-2 µL.

[c]Maximum resolutions were visually estimated from diffraction patterns collected at synchrotron sources (APS – Chicago, IL or SSRL – Stanford University, CA)

[d]Sf9 cells infected with BRIL-PTHR1 baculovirus and grown for 41 hrs were lysed by sonication in presence of 300 nM LA-PTH, a long-acting analog of parathyroid hormone and 2mg/mL iodoacetamide. Isolated membranes were extracted with 1% n-Dodecyl β-D-maltoside (DDM) and 0.1% cholesteryl hemisuccinate. Extractant was incubated with Ni Sepharose 6 Fast Flow (GE Healthcare) and eluted with 250 mM Imidazole after exchanging 1% DDM to 0.1 % Decyl Maltose Neopentyl Glycol (DMNG). Pooled elution fractions were batch bound with anti-FLAG M1 agarose affinity gel **(**Sigma), eluted with 10 mM EDTA and 0.1 mg/mL Flag peptide (Sigma) before being subjected to gel-filtration chromatography (Superose 6, GE Healthcare).

[d]Purified Globin-X from Zebrafish was a generous gift from the lab of Dr. Mark Gladwin.

[e]Transformed BL21(DE3) cells were grown at 37 °C to an optical density of 0.6 at 600 nm, induced with 0.4 mM IPTG for 16 h at 18 °C. Purification was first performed by Ni-NTA chromatography (GE Healthcare) followed by a gel-filtration chromatography (Hi-Load Superdex20016/60, GE Healthcare) equilibrated in 25 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1 mM DTT, 10% glycerol, and 0.02% sodium azide.

Subsequent visual inspection of crystallization drops was achieved using an Olympus SZX16 bright-field microscope. Granular aggregates, as previously defined in (*174*, *175*), and visible crystals that may be used for nanoseeding experiments were assayed for UV tryptophan fluorescence with a Jansi UVEX microscope. UV-positive samples were visualized using the Jan Scientific CrystalDetect software.

5.2.2 Microcrystal fragmentation

Crystals were washed with a total volume of 10 µL mother liquor to remove excess protein from the crystallization drop (this is highly recommended to prevent bubble formation during vortexing). Initial fragmentation of microcrystals using a 3 mm Teflon ball (Hampton Research) and vortexing, revealed minimal and highly irregular crystal fragments (see main text section 3.1 and Figure 23C). To achieve better fragmentation, we utilized smaller diameter (0.1, 0.5 and 1.0 mm) glass beads (Research Products International) or stainless steel beads (Jinan Huawei Industry Manufacturing and Trading Co. Limited) that are routinely used for bacterial or yeast cell lysis. Approximately 25-30 mg of beads were placed inside a 1.5 mL microfuge tube and washed twice with water followed by stabilizing solution. Crystalline material from several crystallization drops (with over 30-50 crystals/drop ranging from 20-100 µm) was added to beads on ice, and diluted with 30 µL stabilizing solution (sufficient to cover the glass beads). Beads and sample were vortexed twice for 10 s, followed by centrifugation at 1,000*g* for 30 s. After centrifugation, the solution was pipetted gently to resuspend crystal fragments and aspirated to a clean 0.5 µL microfuge tube for subsequent experiments. Two microliters of the crystalline solution were evaluated using brightfield and UV microscopy to determine the efficiency of fragmentation (see main text, section **5.2**). Ensuing TEM analysis of crystal fragments or crystal optimization via previously well-established seeding methods (*156, 159*) were performed immediately following the fragmentation protocol. Briefly, 0.3 – 0.4 millimeter loops (Hampton) were used to deposit ~50 nL of nanoseed solution into 1 µL crystallization drops.

5.2.3 Nanocrystal Fragmentation

Conditions containing UV-positive granular aggregates, as previously described above and elsewhere (*175*), were selected for TEM imaging to confirm the presence of nanocrystals. To ensure sufficient material for fragmentation and TEM analysis, crystallization conditions yielding nanocrystals were repeated to generate larger volumes of nanocrystals for fragmentation experiments. Material from 4-12 1 µL crystallization drops (depending on the granular aggregate density) containing nanocrystals were pooled and fragmented as described above with the following caveats: 1) Approximately 8-10 0.5 mm glass beads, or stainless steel beads were used with a total volume limited to 15 µL of stabilizing solution; and 2) samples were vortexed for 5 s once, to avoid sample destruction. Fragmented nanocrystals (heretofore known as nanoseeds, since TEM analysis revealed crystal lattices with nanometer dimensions) were evaluated with bright-field and UV fluorescence microscopy, which was of paramount importance to qualitatively estimate nanoseed concentration for future TEM or seeding experiments.

5.2.4 Fragmented crystal quantification using UV microscopy

Crystal fragments were diluted with stabilizing solution in different ratios, including 1:1, 1:2, 1:5, 1:25 and 1:125 in order to determine seed concentration. A 5µL aliquot from each dilution was placed on a haemocytometer (Hausser Scientific) and covered with a 22 mm glass coverslip (Hampton Research). Bright-field and UV-microscopy images were acquired for 0.1 and 4 s, respectively, with a Jansi UVEX microscope. Crystal images collected under nominal 15x magnification were recorded using the Jan Scientific CrystalDetect software. Manual quantification of crystal fragments was performed using the Cell Counter ImageJ software plugin

and the protocol for counting crystal fragments followed standard cell counting procedures (*176*).

5.2.5 Transmission electron microscopy experiments (TEM)

All crystals generated from proteins listed in Table 4 were subjected to analysis by negative-stain TEM. Approximately 5 μL of nanoseeds were applied to 400 square mesh copper grids covered with a thin continuous carbon film (Electron Microscopy Sciences) and made hydrophilic by glow discharge (EmiTech) for 1 min at 25 mV under atmospheric conditions. Samples were incubated on the grid for 30 sec before blotting and staining with a 2% uranyl acetate solution. Grids were mounted on a standard room-temperature specimen holder and inserted into an FEI Tecnai T12 microscope (FEI, Hillsboro OR) operating at 120 keV. Images were collected on a Gatan Ulstrascan 1000 CCD camera (Gatan, Pleasanton CA), and FFTs were generated using Digital Micrograph version 3.9.4 software (Gatan Software Team, Pleasanton, CA). Preparation of sample grids, data collection, and evaluation of crystal images can be achieved in ~2 hours.

5.3 RESULTS

5.3.1 Crystal fragmentation analysis using UV-microscopy and TEM

Generally, crystal sizes resulting from crystallization screens are too large (Figure 23A) for direct observation of lattices by TEM, which is limited to < 3 µm along the dimension perpendicular to the electron beam (*172*).

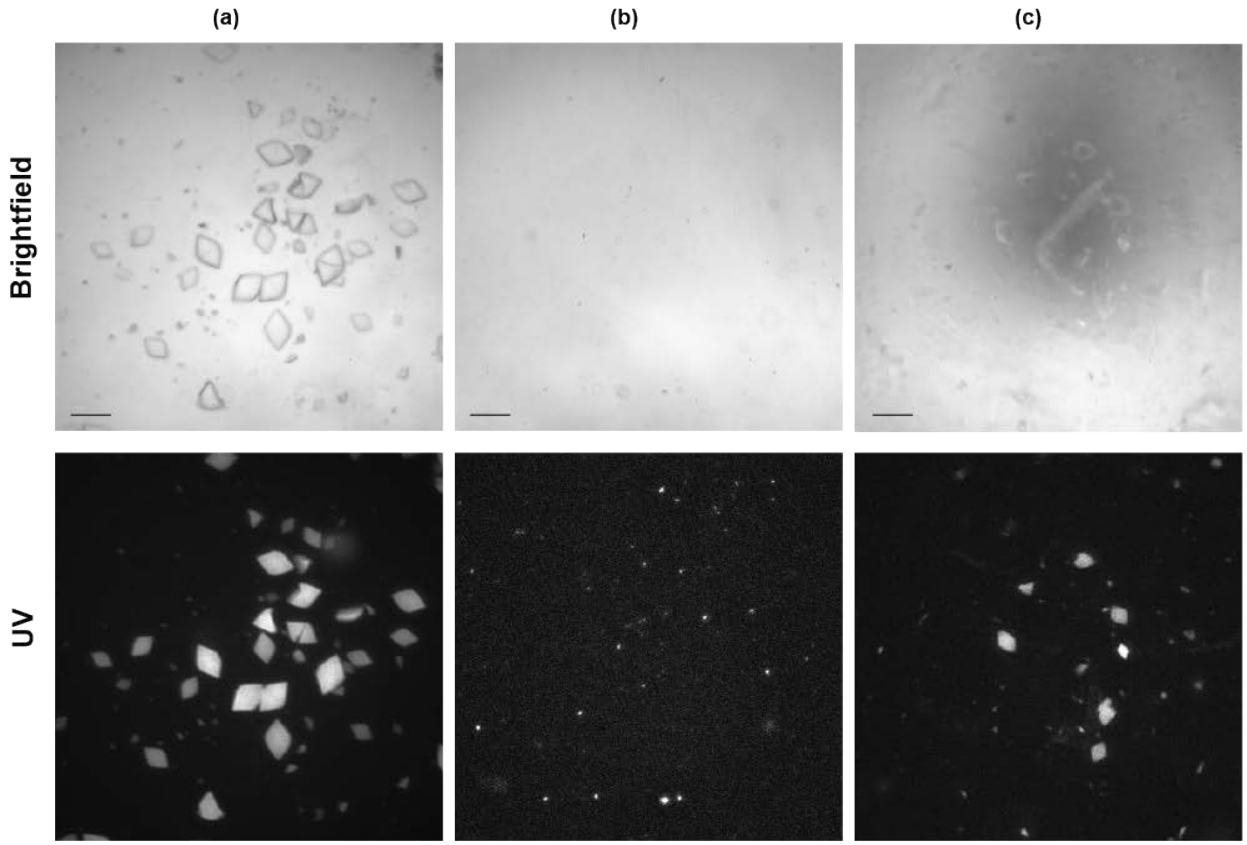

Figure 23. Crystal fragmentation analysis using UV-microscopy and TEM

(a) *Intact crystals used for fragmentation experiments.* ***(b)*** *Crystals fragmented using 0.5 mm beads (see main text section 2.2) which resulted in homogeneously sized fragments measuring less than 5 µm in size.* ***(c)*** *Crystal fragments resulting from use of a single 3 mm Teflon bead according to the manufacture's protocol (Hampton). Fragments were 2-5 times larger than in (b). (Scale bars: 100 µM).*

Thus, a fragmentation protocol was established to generate crystal sizes capable of TEM analysis (see section 5.1.2 and Figure 23). Initially, we tested the effects of various glass or stainless steel bead diameters on crystal sizes after fragmentation. Use of 0.5 mm and 1.0 mm beads resulted in homogenous populations of crystal fragments of low-micrometer sizes (Figure 23B). In contrast, the standard 3.0 mm bead yielded inhomogeneous fragmentation with large crystal sizes still present in the solution (Figure 23C), while no UV-detectable crystals were observed with 0.1 mm beads. While no differences in crystal fragmentation were observed for glass or stainless steel beads, employing stainless steel beads may allow a more efficient recovery of nanoseed slurry, as beads can be readily removed via a magnet.

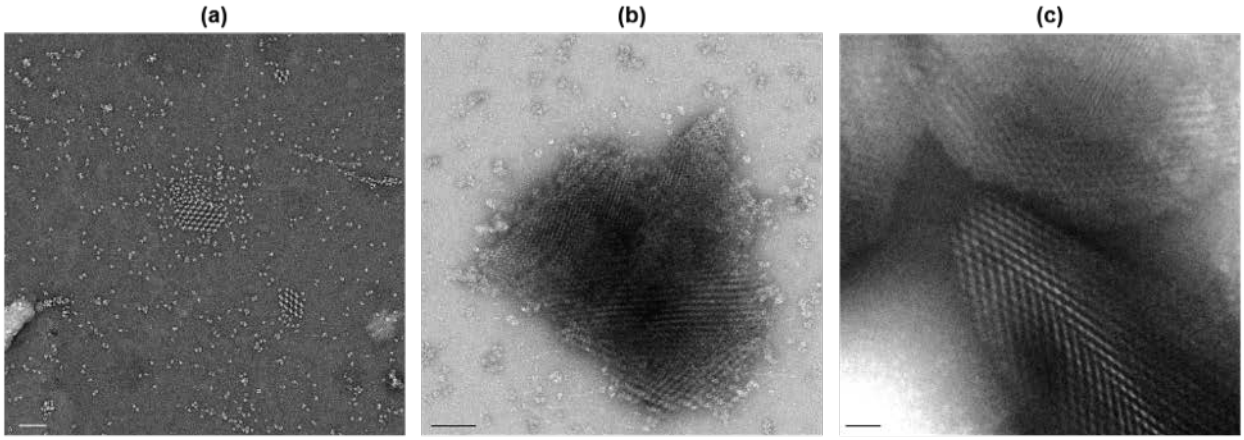

Figure 24. Crystal fragments of Pol-GFP

*Comparison of Pol II-GFP crystal plates crushed with either (**a**) 0.5 or (**b,c**) 1.0 mm beads illustrating the effect of bead size on nanoseed size. Using the same starting material and vortexing time (5 s), crystals in (**a**) were almost completely destroyed with the use of 0.5 mm beads. Crystal morphologies such as plates or needles are highly susceptible to bead disruption and milder protocols should be attempted. (Scale bars: (a) 20 nm; (b) 20 nm; (c) 50 nm).*

After the presence of crystals was initially confirmed with optical or UV microscopy, to

ensure a sufficient concentration for adequate visualization, we utilized negative-stain TEM to analyse our fragmentation protocol with 0.5 and 1 mm beads. Crystal fragments typically ranged from 50 nanometers to a few micrometers in length, and were composed of single to multiple layers (Figure 24). Interestingly, the use of a mild disruption protocol with 0.5 mm beads resulted in the almost complete obliteration of thin crystal plates of Pol-GFP (Figure 24A). Adjusting the protocol to use slightly larger beads was necessary to obtain useful crystal fragments for TEM analysis and crystallization experiments (Figure 24B,C). This suggest that bead size, quantity, and vortexing time are all contributing factors that affect crystal fragment size, as fewer beads with larger diameters and shorter vortexing times yielded larger crystal fragments, and vice versa. However, optimization of these parameters may be necessary depending on crystal type, morphology, and size.

5.3.2 Evaluation of crystal quality using TEM

Since crystalline lattices were clearly visualized with TEM, calculating Fourier transforms from the images allowed qualitative evaluation of crystal lattices (Bragg spots) (Figure 25 and Table 4). In addition, it was possible to establish a correlation between the order of Bragg spots detected and the quality of subsequent X-ray diffraction data. In general, lattices with three or higher order Bragg spots showed diffraction at the synchrotron, whereas crystal fragments with fewer orders of Bragg spots show poor or no diffraction (Figure 26).

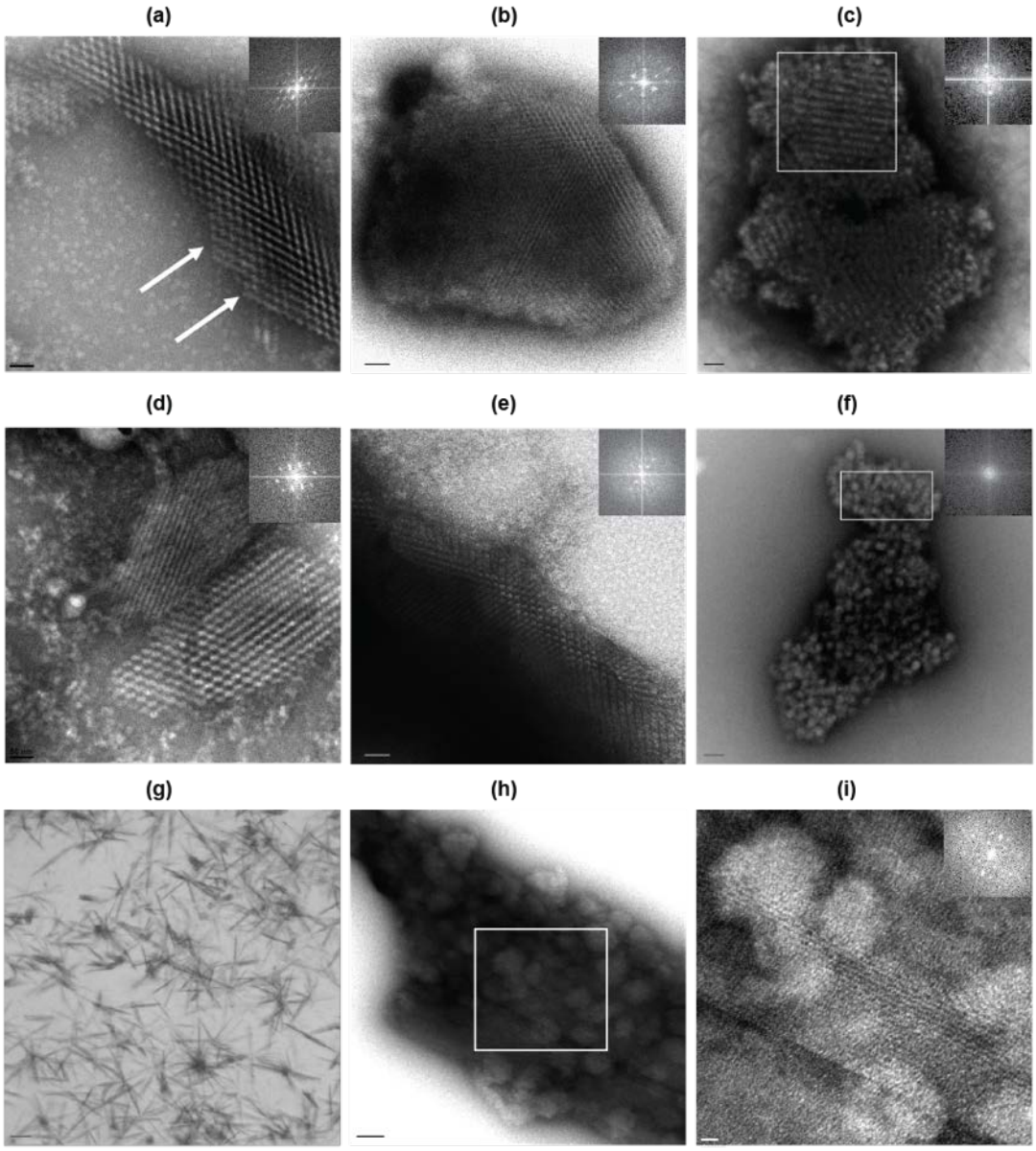

Figure 25. TEM images of fragmented crystals

(a) *and* ***(d)*** *Pol II. Fracture planes (arrows) and individual Pol II particles can be observed.* ***(b)*** *and* ***(e)*** *DDB1-DCAF-UNG2-Vpr crystal fragments.* ***(c)*** *and* ***(f)*** *Sindbis virus crystals showing partially ordered (**c**) and disordered lattices along two different directions.* ***(g)*** *Crystallization of the DDB1-DCAF1 complex showed multi nucleation that was resistant to conventional optimization protocols. (**h,i**) TEM images showing protein aggregation and nanocrystal nuclei growing next to crystal fragments with well-ordered lattices. Use of detergents to decrease protein aggregation resulted in single crystals with improved diffraction to 3.5 Å (see Fig. S3c,d). (Scale bars: (a) 50 nm; (b) 200 nm; (c) 50 nm; (d) 50 nm; (e) 200 nm; (f) 50 nm; (g) 50 µm; (h) 50 nm; and (i) 20 nm).*

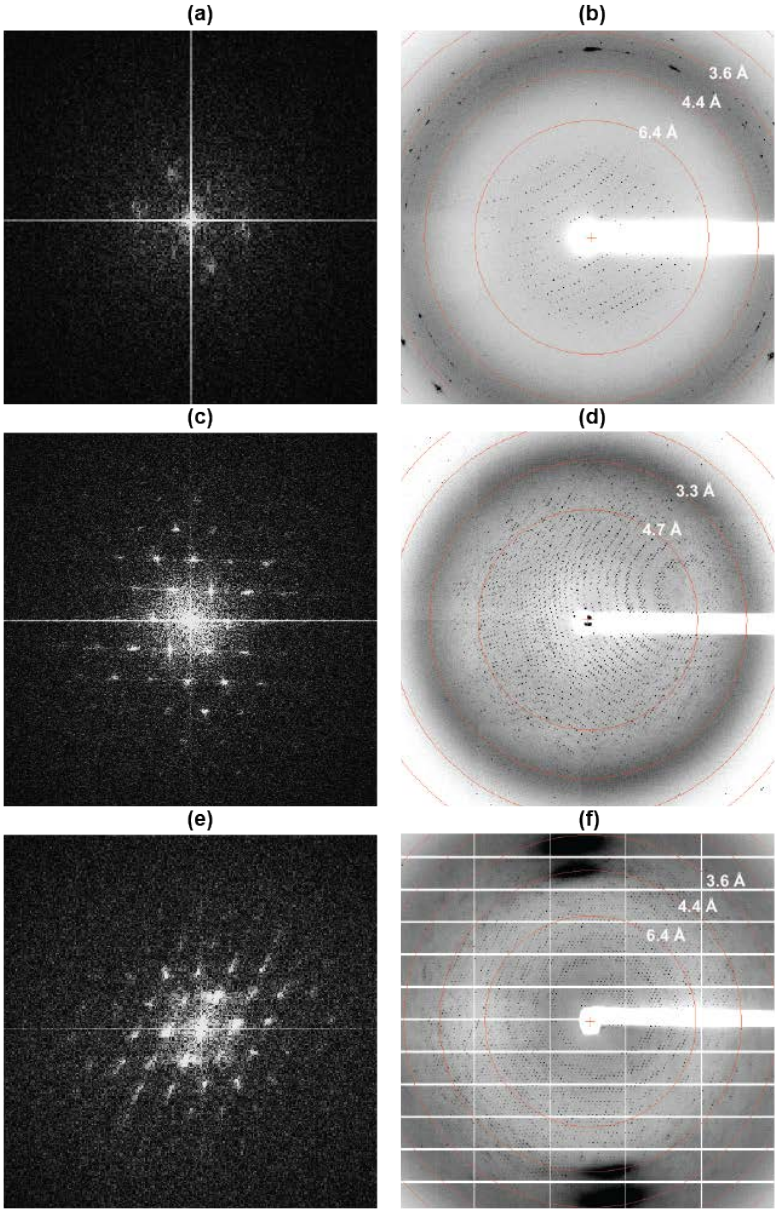

Figure 26. Correlating Fourier analysis of TEM images with single crystal x-ray diffraction

Fourier transforms of lattices (left panels) positively correlate with diffraction images observed at synchrotron sources (right panels), with higher order Bragg spots being a good indicator of higher resolution diffraction. ***(a,b)*** *Rtf1* ***(c,d)*** *E. coli dGTPase, post- dehydration protocols (see Table 4),* ***(e,f)*** *wild-type RNA Polymerase II.*

For example, TEM analysis of the multi protein complex comprising human damage-specific DNA binding Protein 1 (DDB1), DDB1-cullin 4 associated factor 1 (DCAF1) uracil-

DNA-glycosylase (UNG2) and HIV VPR (DDB1-DCAF-UNG2-Vpr) crystals revealed high quality lattices (Figure 25B,E), whereas initial screening of such crystals showed poor to very weak diffraction at the synchrotron (Figure 27A). The presence of well-ordered lattices led us to believe that these crystals could potentially diffract to a higher resolution, which was indeed the case after optimizing cryo-protectant conditions (Figure 27B).

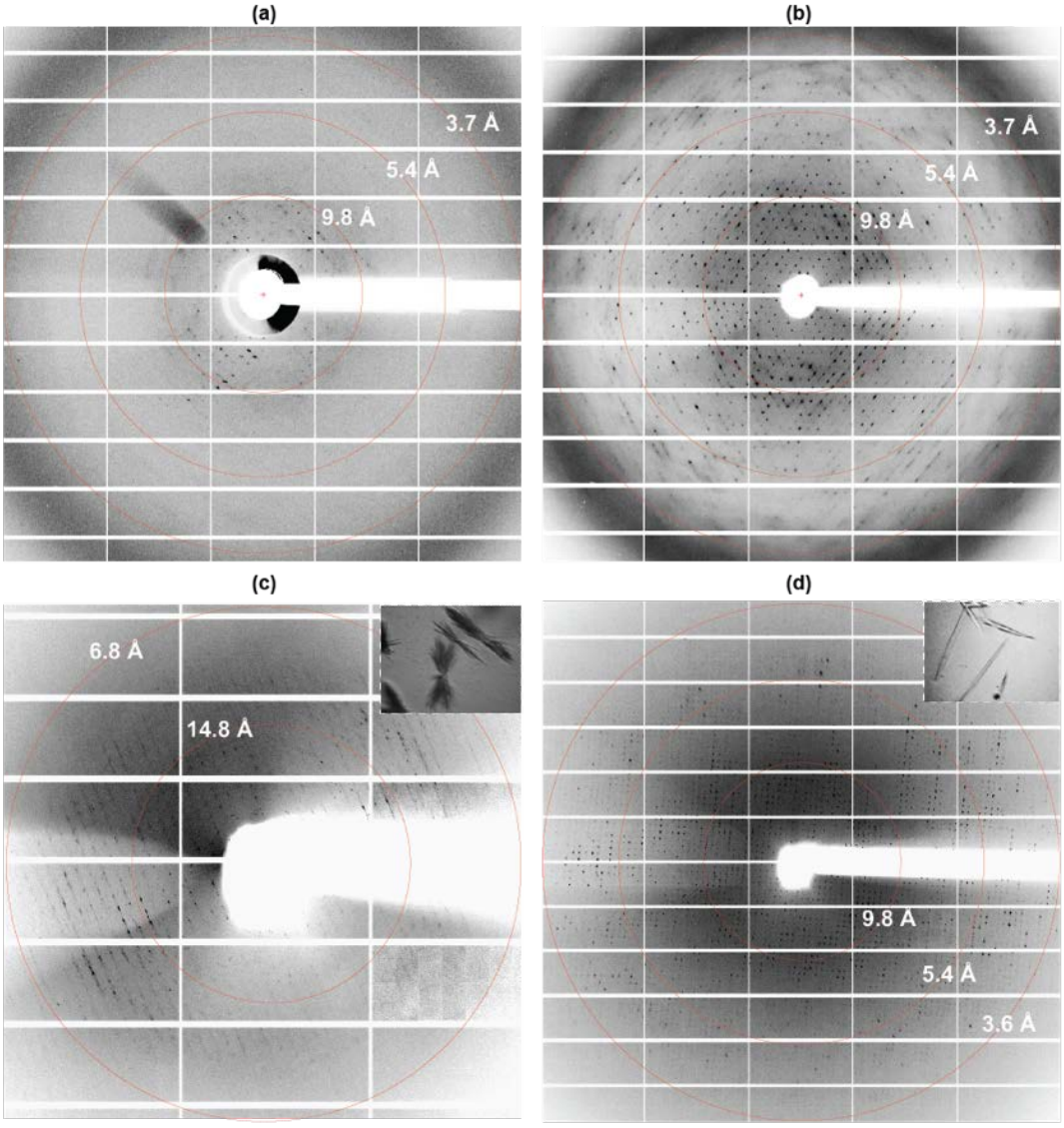

Figure 27. Optimizing x-ray diffraction

(a) *For UNG2-Vpr-DCAF1-DDB1 crystals, characteristic diffraction images of crystals in similar cryo-conditions as previously published structures revealed poor diffraction to ~9 Å. After TEM revealed lattices (see Figure 25)*

showed high-ordered Bragg spots, ***(b)*** *changes in cryo-protectant resulted in improved diffraction to ~3.7 Å.* ***(c)*** *Similarly, DCAF1-DDB1 crystal clusters (inset) diffracted to ~6 Å, but upon the addition of additives* ***(d)*** *single crystal (inset) diffraction improved to beyond 3.6 Å. As seen in Figure 25, the presence of well-ordered lattices for these crystals also suggested improvements in crystallization could result in improved diffraction.*

In contrast, poor x-ray diffraction of partially ordered lattices with low quality Fourier transforms were observed for both Sindbis virus (Figure 25C,F) and the parathyroid hormone receptor (Figure 28), a cell surface G protein-coupled receptor (GPCR) that has proven challenging to crystallize. Taken together with all other protein samples tested (Table 4), these results suggest that crystallization conditions that yield higher order Bragg spots should be pursued to obtain high-resolution structural information.

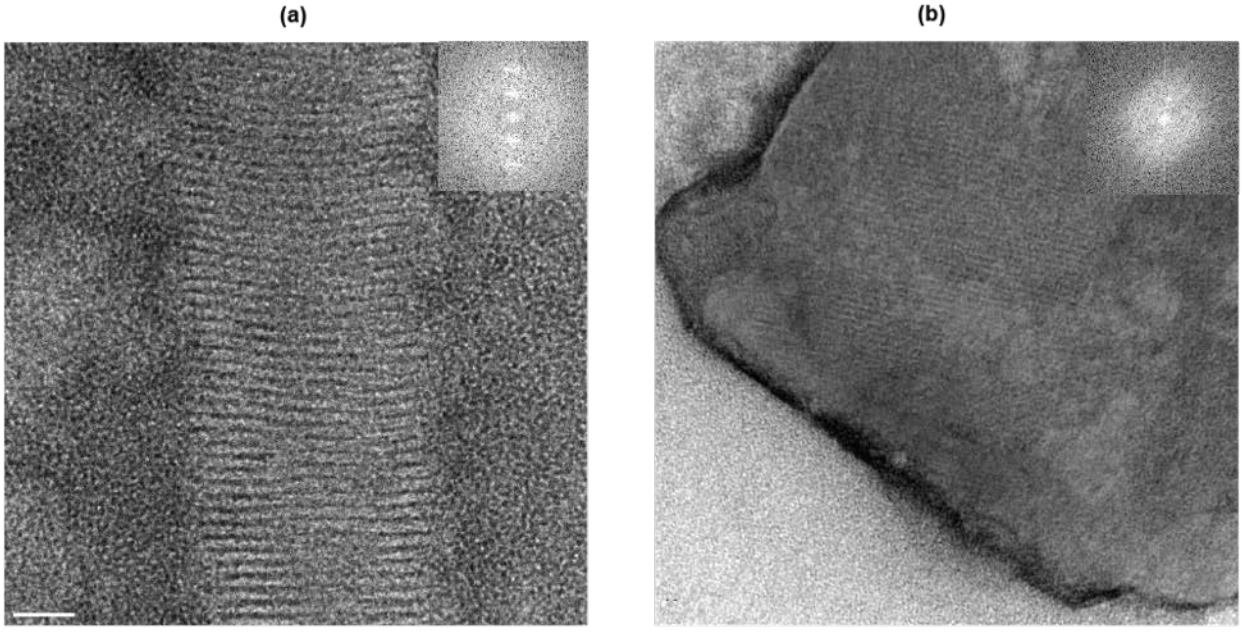

Figure 28. TEM analysis of parathyroid hormone receptor (PTHR) crystals

(a) *Wild- type PTHR1 and* ***(b)*** *Bril-PTHR1 show poor quality lattices. Large crystals of this target have not been reported, but fragmentation of granular aggregates from crystallography experiments showed nanoseed lattices that could be detected by TEM. These crystals may be starting material for matrix-microseeding. (Scale bars: 50 nm).*

In addition to lattice evaluation, TEM provided clues on sample quality and the crystallization process itself. In some instances, crystal lattices were surrounded by multiple nanocrystal-nuclei (of few nanometer sizes) and protein aggregates (Figure 25G-I and Figure 29A-C). In cases where high quality lattices were observed, we tested whether the addition of additives to decrease protein aggregates may improve crystal diffraction. Our results for crystals of the DCAF-DDB1-Vpr complex indicate that such improvements are achievable (Figure 27C,D). Moreover, TEM revealed several cases of protein aggregation, filaments and nanocrystal-nuclei from UV-positive granular aggregates across different crystallization conditions and samples (Figure 29D-F). The capability of TEM to differentiate crystalline material from these other particles is ideal for eliminating false positives that potentially stifle crystal optimization.

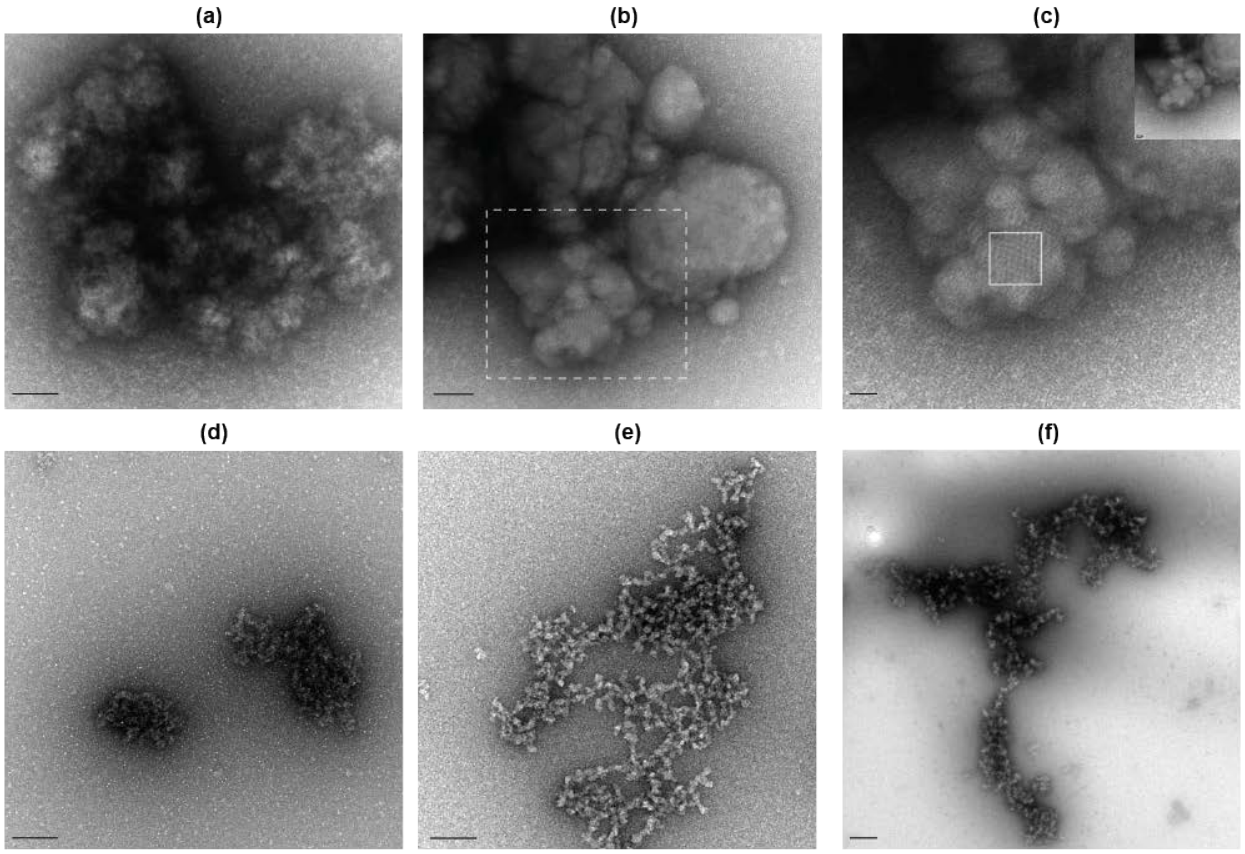

Figure 29. Examples of TEM observed protein aggregates common amongst nanocrystals

(a-c) *The presence of protein aggregates and nanocrystal nuclei among Globin-X crystals compromise small areas of partially ordered lattices (FFT of boxed area, panel (c)). Aggregates found in drops with* ***(d)*** *Pol II-GFP,* ***(e)*** *Wild-type PTHR1 and* ***(f)*** *Bril-PTHR crystals. (Scale bars: (a) 100 nm; (b) 50 nm; (c) 20 nm; (d-f) 50 nm).*

Next, we tested whether anisotropic diffraction could be detected using TEM for crystals where X-ray diffraction data showed prominent anisotropy. Anisotropic X-ray diffraction occurs when the resolution of the diffraction spots varies along crystal axes, and severe anisotropic diffraction can have significant impact on a dataset by limiting its overall resolution. Our results from three different Pol II protein complexes showed anisotropic Bragg spots by TEM, as well as anisotropic X-ray diffraction (Figure 30).

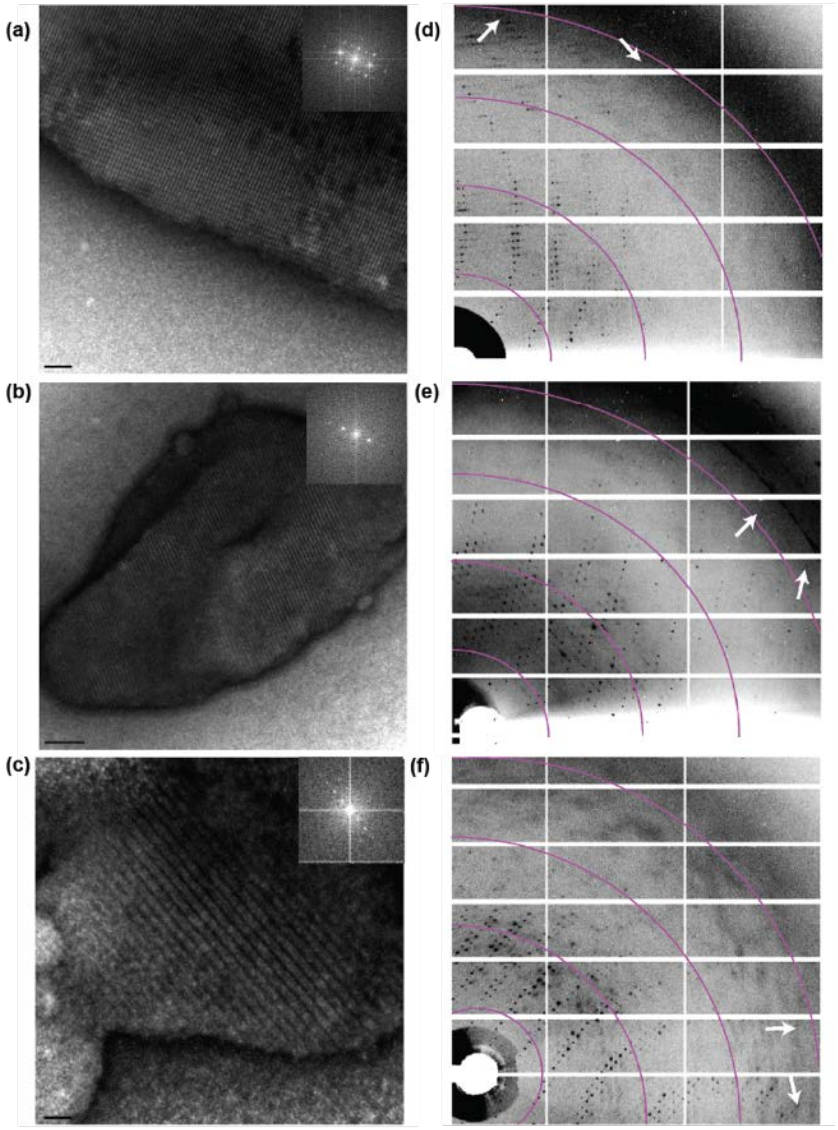

Figure 30. FFT calculation of crystal lattices to predict anisotropy

Anisotropic Bragg spots (left panel-inset) with corresponding anisotropic X-ray diffraction (right panel). ***(a)*** *Pol II-GFP,* ***(b)*** *TFIIF, and* ***(c)*** *Pol II SPT4/5 complex (Scale bars: (a) 100 nm; (b) 200 nm; (c) 50 nm. Diffraction rings (resolution in Å): (d) 20, 9, 6, 4.25; (e) 19, 8, 5.5, 4; (f) 19, 9, 5.5, 3.8.*

Moreover, for one such complex, where different orientations of the crystal lattice were observed by TEM, we could visually align the crystal packing revealed by TEM with the crystal packing after solving the structure using molecular replacement (Figure 31A,B & D,E). TEM

images showed imperfections of the lattice along the direction of anisotropic diffraction due to the presence of scarce crystal contacts (Figure 31 C,F – arrows).

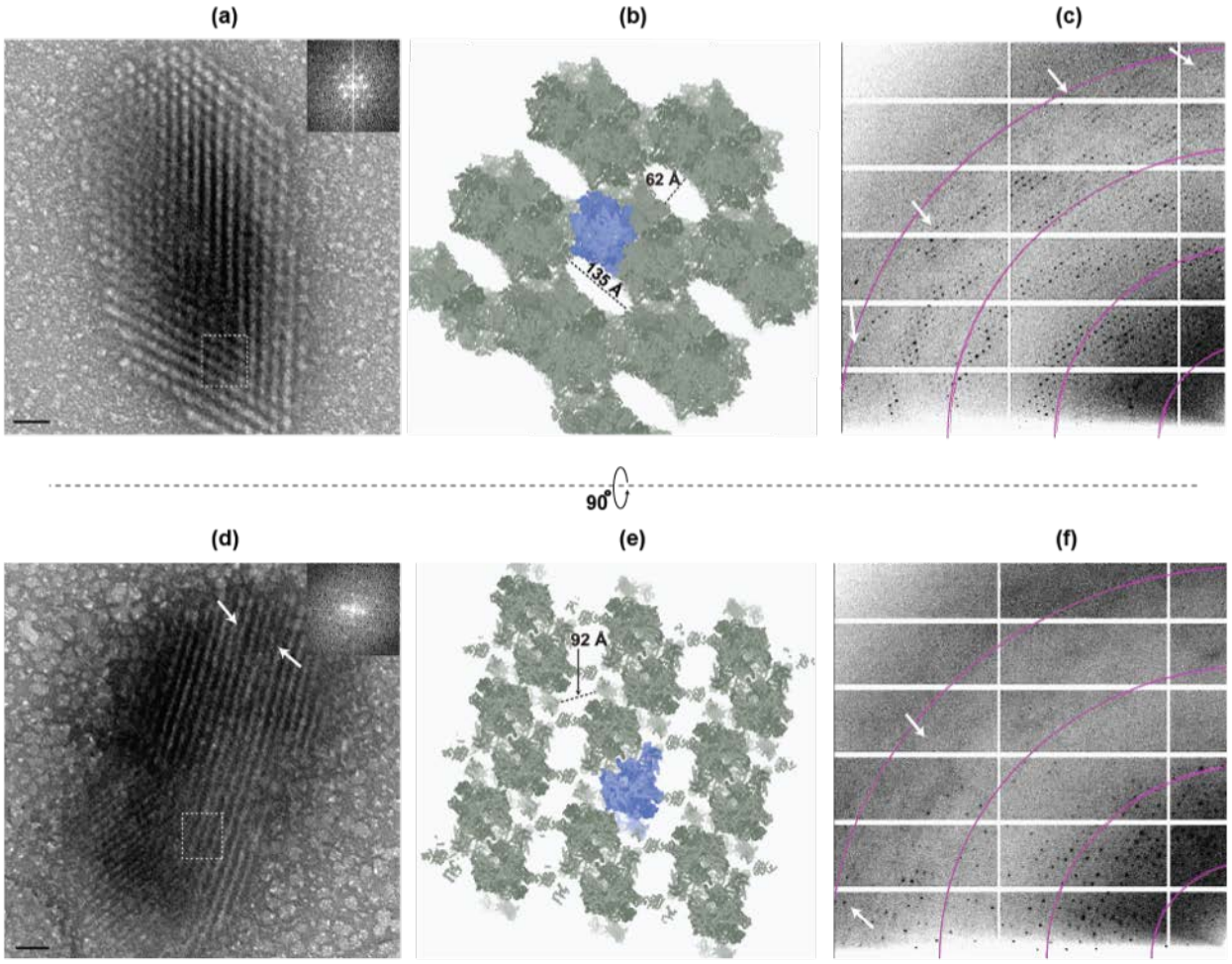

Figure 31. Crystal fragments of Pol II-GFP complex show isotropic and anisotropic Bragg spots along two different orientations

(a-c) *isotropic profile,* ***(d-f)*** *anisotropic profile. Lattice defects can be observed in (d) (arrows).* ***(b)*** *and* ***(e)*** *crystal packing in the* $C222_1$ *space group (cell parameters A=221 Å, B=394 Å, C=211 Å) illustrating solvent channel dimensions (indicated in angstroms). The preliminary model was generated from low resolution data using molecular replacement* (*177*)*, and rigid-body refinement in Buster* (*105*)*. Defects in crystal packing (****d****, arrows) and the large solvent channels (****e****) could explain poor X-ray diffraction along the anisotropic direction in (****f****). (Scale bars: (a) 100 nm; (d) 50 nm. Diffraction rings: (c) and (f) 15, 7.6, 5.1 and 3.7 Å).*

Another example of successfully using TEM to detect crystal lattice defects allowed us to explain uneven diffraction during crystal rastering at the synchrotron for Pol II – Spt4/5 crystals (Figure 32A). For such crystals, a protocol for X-ray rastering was determined after discovering areas of inconsistent diffraction along the crystals as described previously (*178*). TEM images of crystal fragments revealed the presence of well-ordered regions interspaced with patches of blurred lattices elements (Figure 32B). These lattice features were not typically observed for other samples.

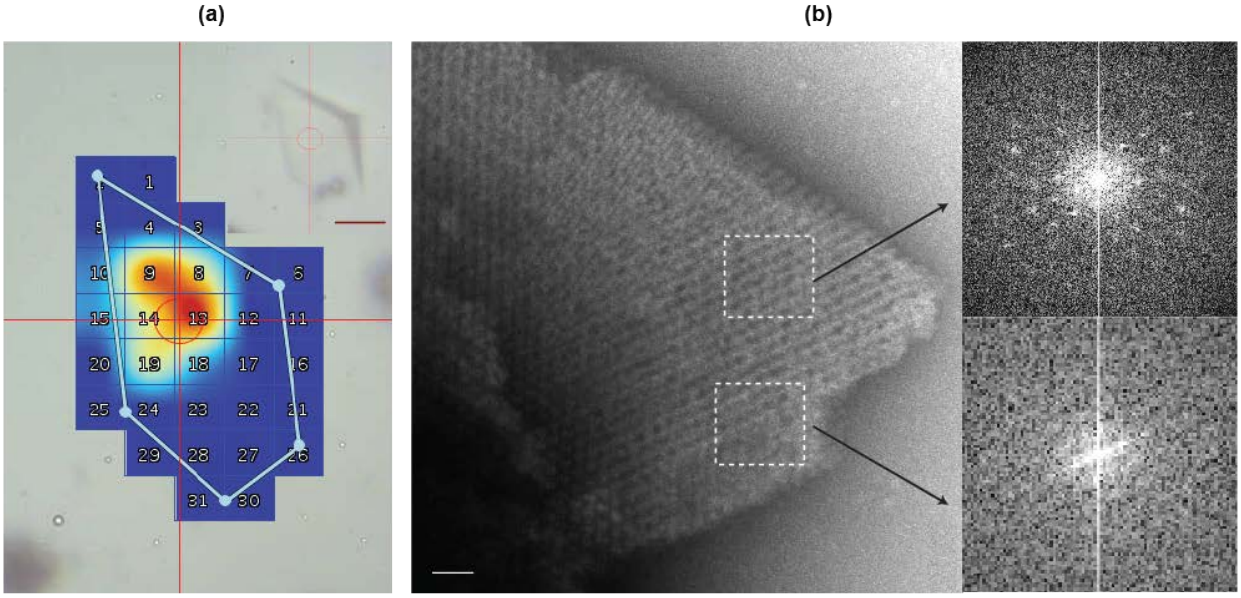

Figure 32. Crystal fragments of a transcribing Pol II – Spt4/5 complex

(a) *X-ray rastering using a highly attenuated beam showing diffracting (red) and non-diffracting (blue) areas within the crystal.* ***(b)*** *TEM images showing well-ordered areas interspersed with poorly-ordered areas as evidenced by calculation of Bragg spots. (Scale bars: (a) 20 μm;(b) 50 nm).*

Furthermore, lattice inspection using TEM is useful for qualitative estimate of solvent content, in particular for crystals of large molecular weight proteins (Figure 33). Such samples, where relatively large solvent channels are observed, may benefit from the use of dehydration

protocols to manipulate crystal solvent content and potentially improve diffraction (*173*). Indeed, the use of post-crystallization dehydration protocols was paramount in the structure determination of RNA Polymerase II crystals (*25*).

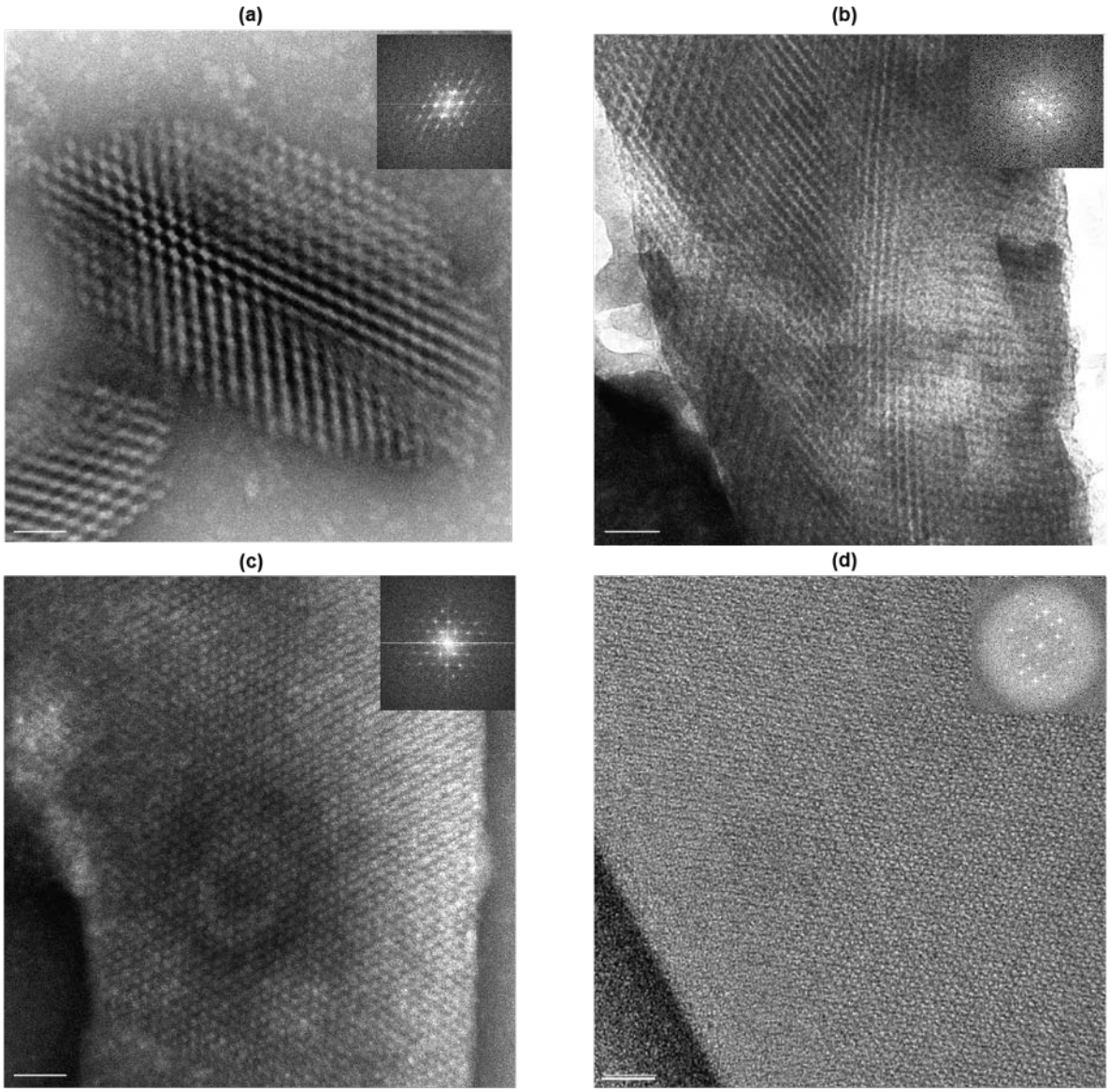

Figure 33. Qualitative evaluation of crystal fragment's solvent content

(a) *Wild type Pol II,* ***(b)*** *Pol II-Spt4/5,* ***(c)*** *dGTPase and* ***(d)*** *H200Q variant of homoprotecatechuate 2,3-dioxygenase. Calculated solvent contents after Matthew's coefficient calculation are 78%, 65%, 58% and 50% respectively. (Scale Bars: 50 nm).*

Thus, we tested whether improvements in lattice quality could be visualized after the dehydration of RNA polymerase II (Δ4/7 Pol II) and *E. coli* dGTPase crystals. Such dehydration protocols resulted in higher resolution diffraction at the synchrotron for the dGTPase sample (Figure 34). TEM analysis of pre- and post-dehydrated crystals revealed improved lattice quality, as indicative by higher-ordered, isotropic Bragg spots for both samples (Fig. 34).

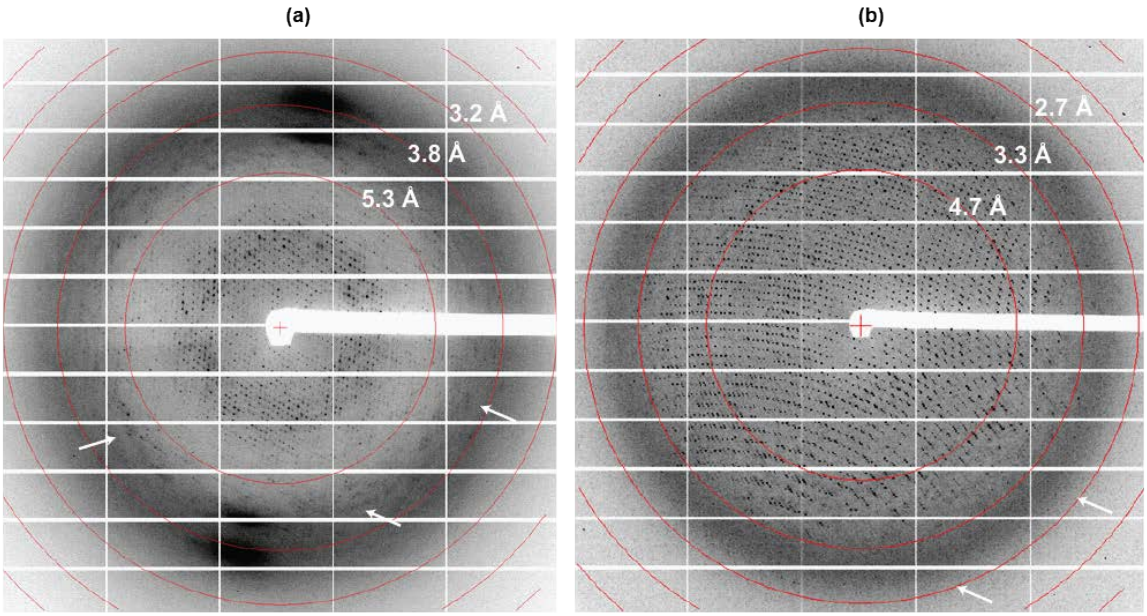

Figure 34. dGTPase dehydratin diffraction comparison

Characteristic diffraction images of dGTPase crystals pre- ***(a)*** *and post-* ***(b)*** *dehydration protocols. Arrows represent maximum observed reflections of ~ 4.3 and 2.9 Å, respectively.*

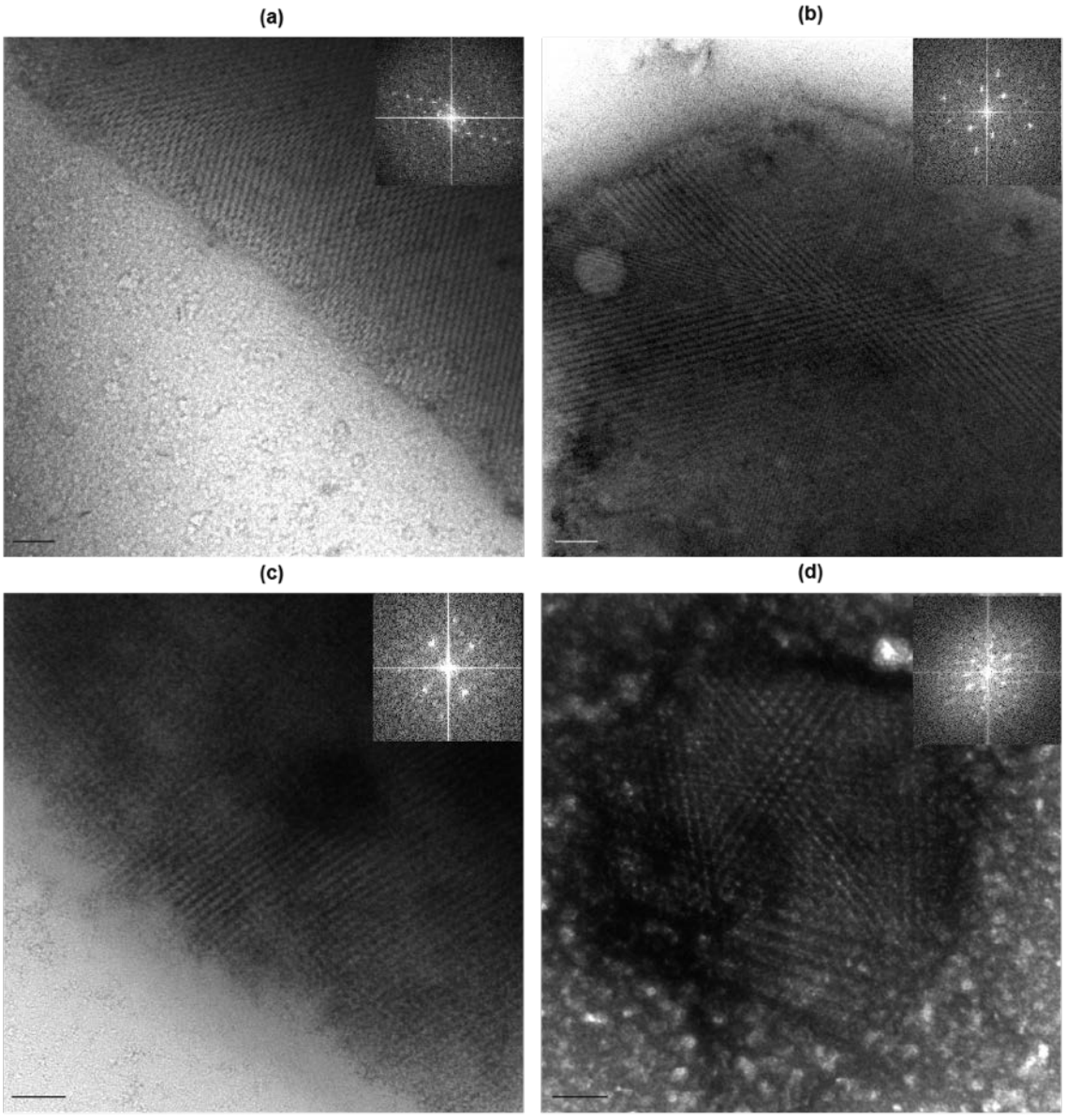

Figure 35. Evidence of lattice quality (order of Bragg spots) improvement upon implementation of crystal dehydration protocols

(a) *Δ4/7 Pol II nanocrystals before dehydration experiments showing anisotropy,* ***(b)*** *crystals undergoing dehydration (protocol adapted from (25)) show isotropic Bragg spots. (**c**) and (**d**) dGTPase crystal fragments pre- and post- dehydration (respectively), showing overall improvement of Bragg spots after dehydration. A decrease from 63% to 58% in solvent content was calculated for the different dGTPase crystal forms. Corresponding X-ray diffraction of crystals (before and after dehydration) are illustrated in Figure 34. (Scale Bars: 50 nm).*

Lastly, since nanocrystals can be commonly found in crystallization drops (*174*), we used TEM to determine whether if it were possible to find nanocrystals with higher ordered lattices among different conditions in a crystallization screen. To this end, APOBEC nanocrystals were first identified using UV-microscopy as described previously and applied to EM grids for visualization. TEM analysis revealed differences in crystal quality from several crystallization conditions where crystalline lattices were observed (Figure 36). Similar to the identification of protein aggregates or filaments from UV-positive crystal drops (Figure 29), these data suggest that utility of TEM for providing the highest quality starting material for subsequent crystal optimization experiments.

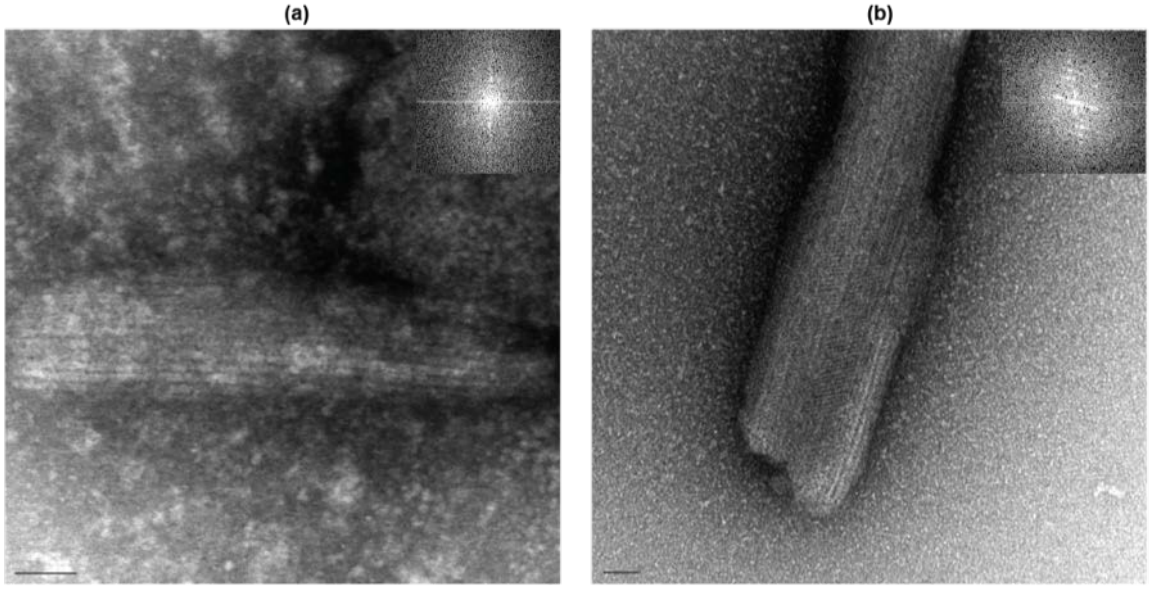

Figure 36. Crystal fragments of Apobec obtained from two different crystallization conditions

*(**a**) 1.5 M Sodium Citrate (**b**) 200 mM Magnesium Acetate + 10% PEG4000 (Scale bars: 50 nm). Crystals in (**a**) show a poorly ordered lattice with no Bragg spots while the crystal lattice in (**b**) shows 2^{nd}-3^{rd} order Bragg spots, which may be indicative of a condition which will produce high resolution crystals.*

3.3 Quantification and nanoseeding experiments of crystal fragments

The presence of well-ordered crystal lattices for various Pol II complexes (as observed by TEM) led us to utilize these crystal fragments for microseeding experiments. Traditional microseeding relies on serial dilutions of seeds used in optimization experiments to control the quality, concentration and size of crystals that develop in the drop (*156*). However, these dilution series traditionally vary in seed concentration with every preparation, and so limiting their reproducibility. Given that the size of crystal fragments were fairly homogeneous, we developed a protocol to quantify the size and number of crystal fragments using UV microscopy and a haemocytometer (see methods section 5.1.4 and Figure 37). Using the nominal 15X magnification, we were able to quantify crystals whose smallest dimension was >2 µm (Figure 37A-C). Not surprisingly, dilution series of detectable crystal fragments by UV-microscopy resulted in characteristic changes in crystal size and number (Figure 39A). Moreover, implementation of a quantification protocol allowed reproducible, high-quality crystal growth for the several Pol II complexes tested, as illustrated by the overall improvement in crystal size, resolution and mosiacity of the Pol II – TFIIB – DNA complex (Figure 38A).

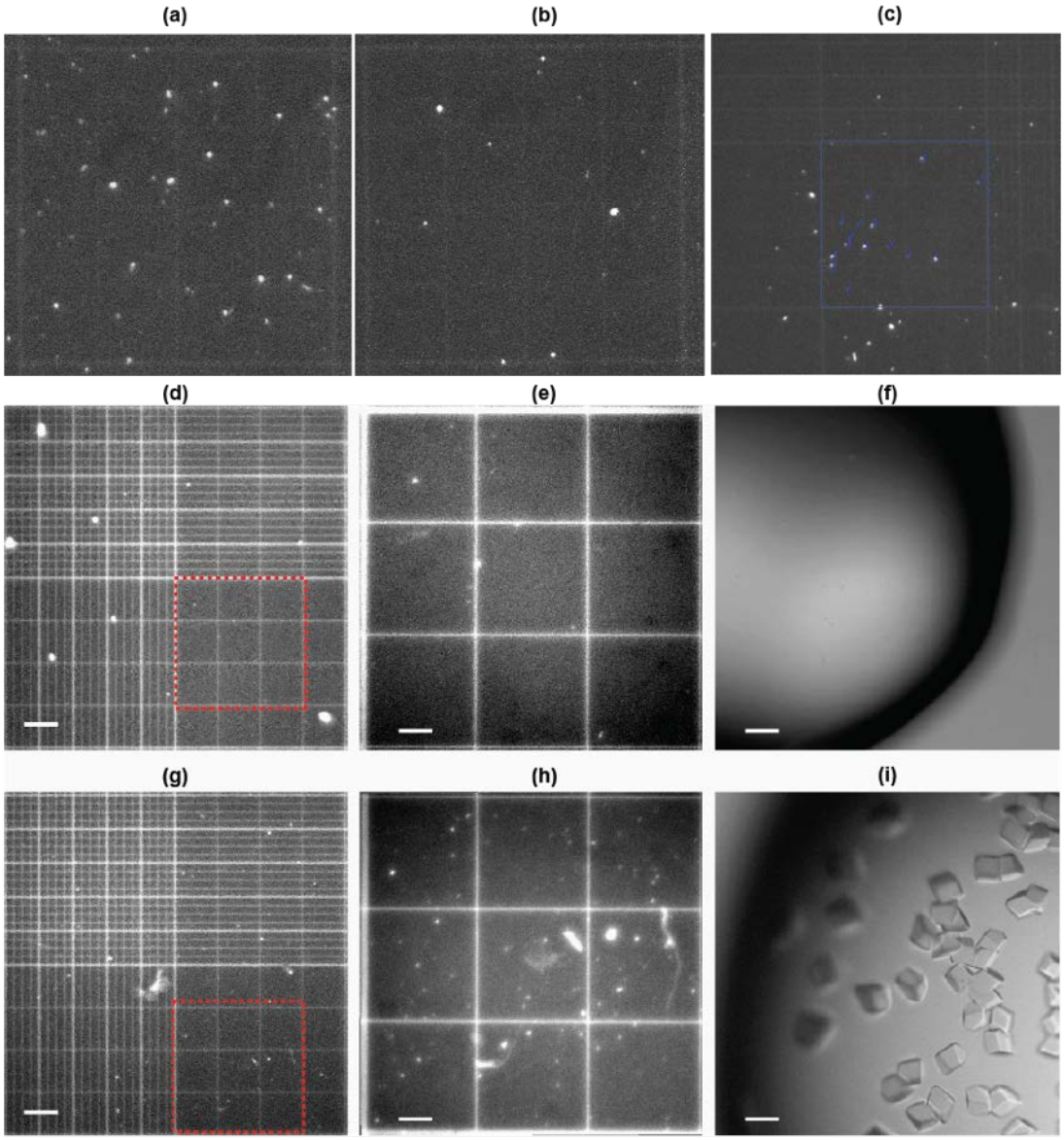

Figure 37. UV tryptophan fluorescence image of seed concentration determination using a haemocytometer

Seeds were counted by pipetting them into the counting chamber of the haemocytometer, covering with a thin UV transparent coverslip, taking a UV fluoresce microscopy image and counting visualized seeds using ImageJ software (see main text section 2.4). ***(a)*** *Undiluted seeds,* ***(b)*** *1:4 dilution of seeds. Total seed counts calculated for the 1x stock taken from* ***(a)*** *and* ***(b)*** *were on average of 330 seeds per µL.* ***(c)*** *Lower magnification (5x) image of haemocytometer with seeds showing the ImageJ counting overlay. In addition, crystals were fragmented with either 0.1 mm* ***(d-f)*** *or 0.5 mm* ***(g-i)*** *beads and quantified with a haemocytometer. Nominal 15x zoomed images revealed significantly higher detectable concentrations of seeds from 0.5 mm bead fragmentation* ***(h)*** *versus 0.1 mm beads* ***(e).*** *Seeding experiments show that over a one week duration, seeds generated from 0.1 mm beads* ***(f)*** *show no crystals while 0.5 mm bead fragments* ***(i)*** *grew crystals (Scale bars: (a,b) 200 µm; (c,d) D 75 µm; (e,f) 100 µm).*

Although crystals with nanometer-sized dimensions were present in our fragmented samples (as visualized by TEM), it was not possible to count these crystals due to the resolution limits of the UV-microscope. Thus, we tested whether nanometer sized nuclei contributed to the observed crystal growth during seeding experiments. The use of 0.1 mm beads for crystal fragmentation of the Pol II – TFIIB – DNA complex resulted in no detectable crystal fragments by UV-microscopy, and seeding experiments using this sample did not yield detectable crystals when compared to drops seeded with the 0.5 mm bead solution over a one week period (Figure 37D-I).

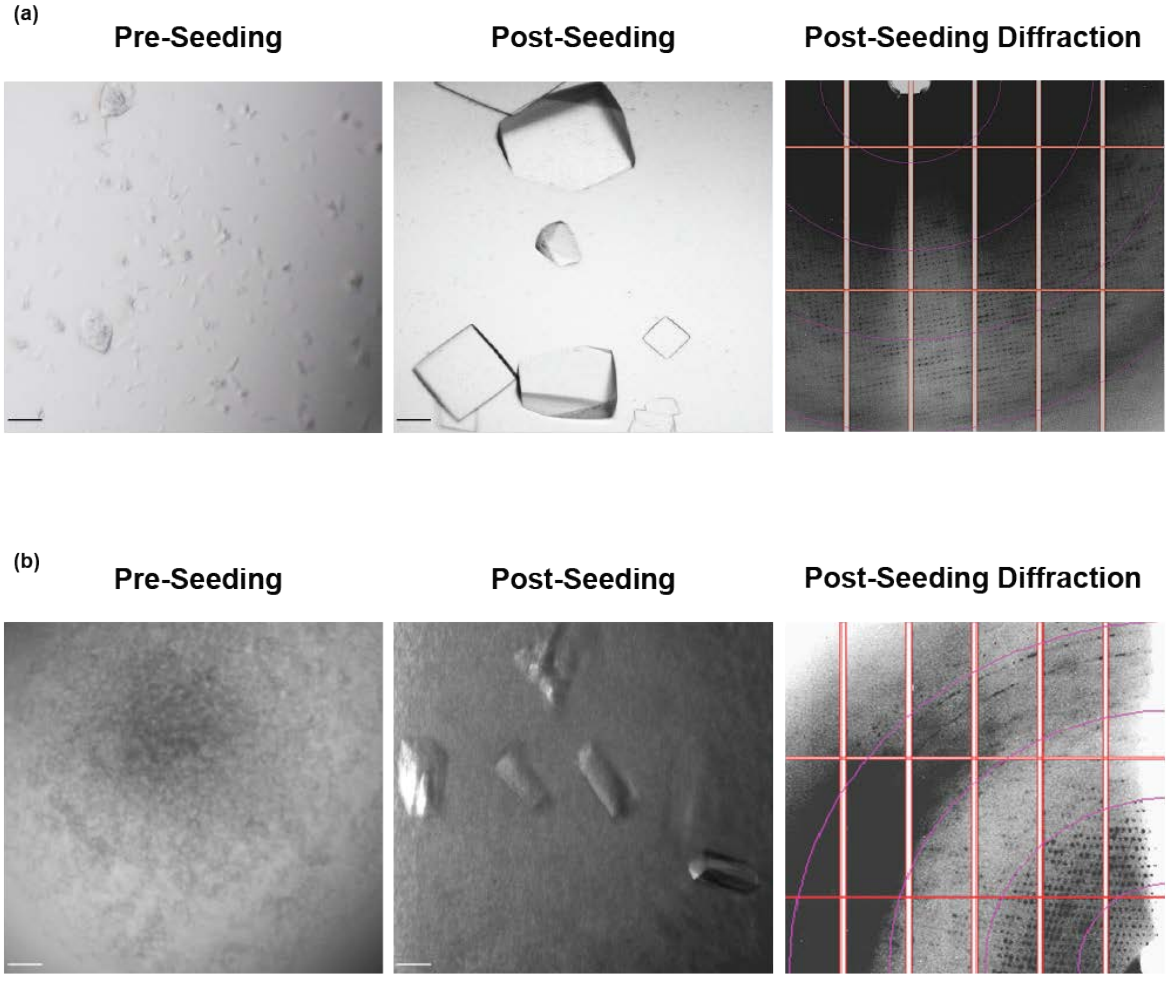

Figure 38. Brightfield images of microseeding experiments with corresponding post- seeded diffraction

(a) *Microcrystals (left panel) of a complex composed of Pol II, TFIIB and a nucleic acid scaffold were fragmented and used for microseeding experiments to obtain larger, well-ordered crystals (middle panel) which showed high-resolution diffraction (right panel).* ***(b)*** *Nanocrystals from granular aggregates (left panel) of a complex between Pol II-CD3Δ were used to generate nanoseeds used in nanoseeding experiments to generate microcrystals (middle panel) which diffracted to high resolution at the synchrotron (right panel). (Scale bars: (a) 100 µm; (b) 50 µm. Diffraction rings: 15, 7.6, 5.1 and 3.7 Å).*

We next tested whether nanocrystals with high quality crystal lattices visualized by TEM could be used for nanoseeding experiments. Given the small size of nanocrystals from granular aggregates, it was not possible to count these samples using UV-microscopy; therefore, no serial dilutions were made since the concentration of nanoseeds (fragmented nanocrystals) was typically low. For the Pol II – CD3Δ complex, ~50 nL of nanoseed solution was deposited into fresh 1 µL crystallization drops as previously described. After three successive rounds of nanoseeding from fragmented nanocrystals, microcrystals identifiable by UV-microscopy were produced, which were then used to generate diffracting crystals (Figure 38B). Overall, these data suggest that 1) TEM identifiable nanocrystals from granular aggregates might lead to an increasing number of conditions that promote crystal growth, 2) UV-detectable crystal fragments provide a better medium to generate larger crystals, and 3) quantification of such fragments leads to highly reproducible results.

3.4 Generation of crystal catalogues for XFEL experiments

To test whether we could generate homogenously sized crystal catalogues for use at CXI or XPP, we utilized quantified seed solutions with varying precipitant-to-sample dilutions for the Pol II – TFIIB – DNA sample. Our results indicate that by varying the precipitant-to-protein concentration ratio and the number of UV-quantified seeds, we could control the size and

quantity of crystals (Figure 39A). Moreover, crystal drops with different ratios of protein-to-precipitant volume, seeded with equal amounts, generated collections of homogeneous crystal sizes (Figure 39B), while such patterns were absent for non-seeded drops. This suggests that generating crystal catalogues consisting of finely tuned crystal sizes may be attainable and reproducible, however additional trials will be needed to establish a protocol.

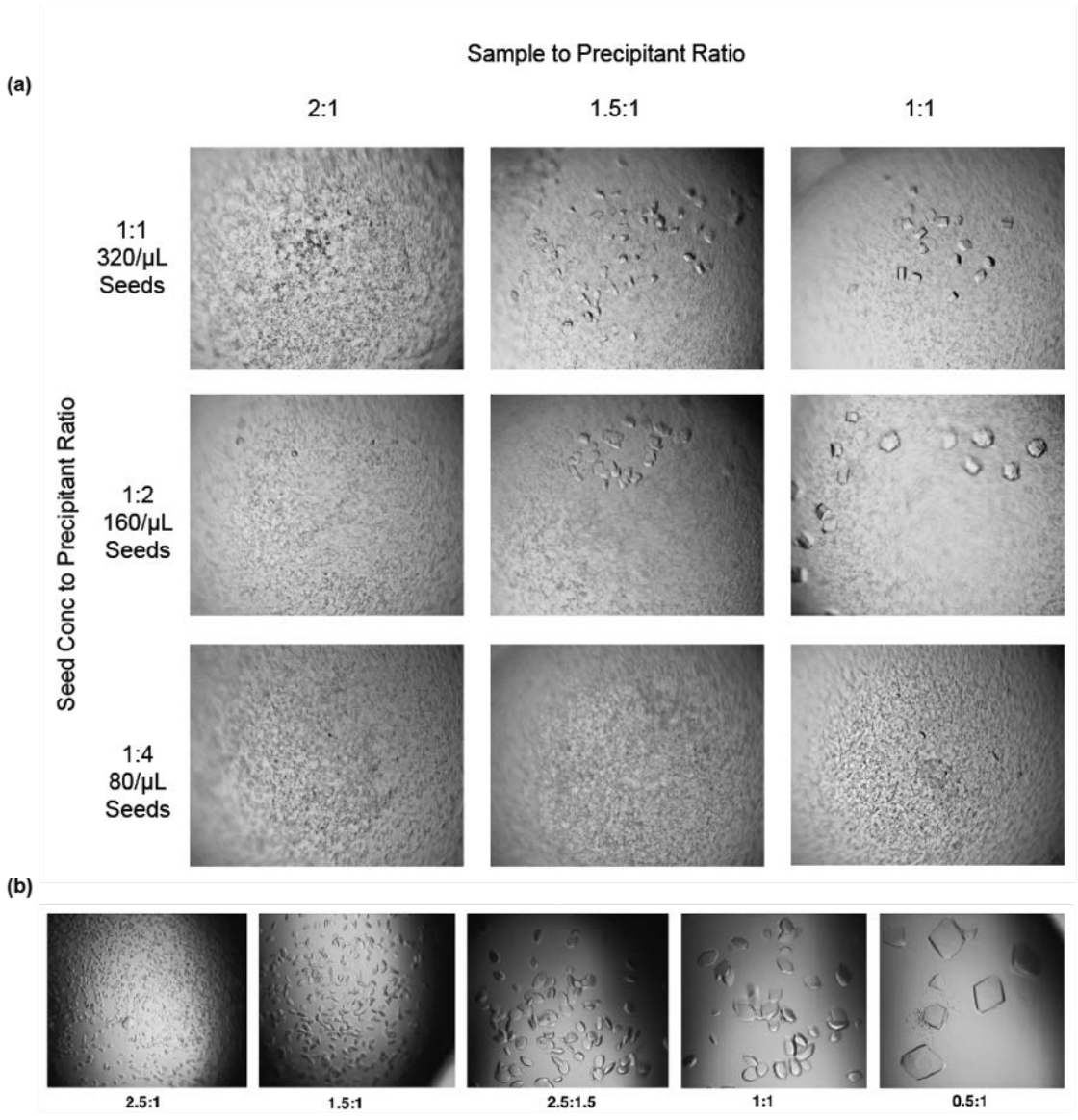

Figure 39. Creating crystal catalogues

Fragmented crystals were prepared using 0.5 mm beads as described in main text section 2.2 ***(a)*** *Comparison of the effect of varying the precipitant to sample ratio with the effect of quantified seed concentration.* ***(b)*** *Refinement of crystal sizes was performed by varying the precipitant to protein ratio. In all cases the same amount crystalline slurry was added to the crystallization drop.*

5.4 DISCUSSION

The process of crystallization remains the most significant challenge in the structure determination of macromolecules by x-ray crystallography. In a collaborative and comprehensive approach, we demonstrated the utility of TEM for the visualization, evaluation, and optimization of the crystallization process. Negative stain TEM analysis of nanocrystals or crystal fragments provides a visual method to evaluate qualitatively the characteristics of the crystal lattice and can be performed in most research institutions. In this work we provide protocols to perform these experiments in a quick and efficient way.

First, TEM images revealed details of the crystal lattice that appeared to be good indicators of the potential diffraction of the crystal at the synchrotron. Given the large number of crystals (from a wide range of "biologically relevant" protein samples) employed in our study, it is feasible to conclude that negative-stain TEM images of fragmented crystals bearing well-ordered lattices (by visual inspection) and at least third order Bragg spots have a significant possibility to diffract at the synchrotron (Table 4 and Figure 26). We provide up to 12 examples where observation of third order Bragg spots on nanocrystals or crystal fragments yielded at least 4 – 5 Å resolution diffraction. We did not observe useful X-ray diffraction for crystals that had disordered lattices of less than 1-2 order Bragg spots. Moreover, this correspondence was also observed for crystals diffracting anisotropically (Figure 30). Thus, the qualitative information gained by analysis of nanocrystals or crystal fragments should allow the experimenter to decide whether a crystallization condition should be explored further and optimized or discarded, saving valuable time during crystallization efforts.

A physical reason for such behavior might be that highly ordered lattices could withstand better the effects of negative staining (i.e. dehydration, the damaging effect of the stain itself and imaging at room temperature). Such behavior may explain the anisotropic diffraction observed for Pol II – GFP crystals in Figure 31. It is possible that the larger number of crystal contacts along the isotropic direction could withstand the effects of crystal dehydration while the direction with fewer contacts collapses. Conversely, we believe that the fragmentation protocol itself has probably little effect on the quality of the lattice itself for two reasons: 1) protocols for fragmentation of crystals for nanocrystallography at free electron lasers have been employed and such crystals diffract at comparable resolution to their larger counterparts (*175*, *179*), and 2) lattice comparison of grown nanocrystals with fragmented microcrystals show no difference (in terms of quality) when visualized with negative-stain TEM (*180*).

Secondly, we demonstrated the utility of TEM to identify crystal pathologies that contribute to poor growth and x-ray diffraction data. Similar to experiments utilizing atomic force microscopy (*181*, *182*), crystal pathologies revealed by TEM included: 1) crystal "contamination" by heavy protein aggregates and micro-crystal nuclei (Figures 25H-I & 29A-C); 2) anisotropic diffraction (Figure 30); and 3) crystal lattice defects (Figures 31D & 32B). Detection of lattice defects in some crystals could point to the presence of samples containing protein contaminants, aggregates or partially proteolyzed protein as well as discrepancies in the stoichiometry of the sample. Similarly, identifying crystals that possess anisotropic diffraction may indicate that steps to improve crystal contacts, such as altering or adding reagents to the crystallization conditions (*183*) or modification of the protein itself, may be advisable. The presence of heavy protein aggregation or nanocrystal nuclei indicates that increased protein solubility or decreased crystal nucleation may be useful to improve diffraction.

Third, TEM enabled a qualitative estimation of crystal solvent content and permitted us to explore the effect of lattice dehydration on crystal diffraction. This application was particularly noteworthy since 1) crystal dehydration protocols have proven very useful in the improvement of X-ray diffraction (*173*, *184*), and 2) negative-staining with uranyl acetate provides very high contrast between solvent channels and biological macromolecules. Furthermore, decreasing the solvent content for Pol II Δ4/7 and dGTPase crystals (13% and 5% decreases as calculated by Matthew's coefficient, respectively) showed improvement in the order of Bragg reflections (Figure 33), which correlated with improved x-ray diffraction (Figure 34). Although our assessment of solvent content was qualitative, TEM images provide a means to accurately estimate crystal solvent content through evaluation of the crystal symmetry observed in FFTs and measured protein particle size, as described elsewhere (*185*). Such information could potentially guide post-crystallization protocols to help improve x-ray diffraction.

Furthermore, our work demonstrates a comprehensive approach to seeding through: 1) TEM identification and evaluation of crystal lattices, and 2) the use of UV-microscopy to quantify seed concentration using conventional cell counting protocols. While the use of granular aggregates have provided suitable starting material for microseeding experiments (*159*), success of applications like MMS will benefit greatly from methods which are capable of detecting crystalline material. Semi-automated applications such as SONNIC and UV fluorescence techniques are capable of visualizing nano-meter sized crystals, but they lack the ability to provide information on the quality of the starting seeding material. For instance, crystallization experiments of Pol II-TFIIF, PTHR1 and Sindbus virus yielded nano- and microcrystals for all samples; however TEM analysis allowed identification of lattice defects (Figures 25C,F, 27B and 28), which suggested that these crystals were unlikely to succeed as seeding nuclei.

In contrast, nanocrystals identified by TEM from a crystallization screen for the Pol II-CD3Δ complex which showed 2nd order Bragg reflections (*168*), produced diffraction quality crystals after multiple rounds of iterative microseeding (Figure 38). Such cases as these suggest that exploring multiple conditions within a screen may lead to higher quality starting material for microseeding experiments to optimize crystal growth. Ideally, an analytical procedure involving identification by SONICC or UV fluorescence of protein nano-crystals, followed by scoring through TEM analysis would greatly aid the crystallographer. Indeed, such a procedure was utilized after UV fluorescence identification of APOBEC crystals from a matrix crystallization screen. TEM analysis of crystals generated from different crystallization conditions, demonstrated the utility of TEM in distinguishing ordered crystal lattices from crystals with lattice defects (Figure 36). In our experiments, a direct correlation between seed quality and the appearance of high quality crystals in seeded drops was easily established and corroborated by X-ray diffraction data (Figure 38).

Finally, critical to our TEM analysis was the development of a crystal fragmentation protocol that produced homogenously sized crystal fragments. While fragmentation protocols have been used to generate homogenously sized crystals that diffract to high-resolutions at XFEL sources (*179*), homogeneous sample preparation for both XFEL and microED techniques is an active research area (*180, 186-188*). Quantification of crystal fragments by UV microscopy, allowed for highly reproducible seeding experiments (Figure 38A) and generation of homogenously sized crystal catalogues through manipulation of crystallization conditions (Figure 39). Application of this protocol to batch crystallization methods, like those commonly used for sample preparation in SFX or microED experiments, would likely generate the high quantity of uniform crystals required for these experiments. Moreover, such experiments can be

complimented by TEM analysis to ensure sample homogeneity, monodispersity, and quality prior to conducting experiments, thus increasing the chances of success.

Overall, we established a reproducible protocol using TEM to qualitatively visualize fragmented crystals, with nano- to micro-meter dimensions, to study details of the crystallization process. TEM analysis allows for the direct comparison of the quality of crystal lattices among different crystallization conditions, thus guiding the experimenter's selection of crystallization conditions that promote well-ordered crystal growth and high-resolution diffraction. In order to probe the real limits of diffraction of such nanocrystals or crystal fragments, TEM experiments using cryo methods and the use of special detectors to probe electron diffraction would be needed. Overall, we believe that the screening methods described here provide a rationalized approach to meet the challenges associated with obtaining suitable crystals for macromolecular crystallography experiments at both synchrotron and XFEL sources.

5.5 ACKNOWLEDGEMENTS

Portions of this research were carried out at the Linac Coherent Light Source (LCLS), a National User Facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. We thank Sebastien Boutet, Marc Messerschmidt, Daniel DePonte and Garth Williams of LCLS, and Robert L. Shoeman and Sabine Botha of the Max Plank Institute for Medical Research for support during data collection at coherent X-ray imaging (CXI). The CXI instrument was funded through the LCLS Ultrafast Science Instruments (LUSI) project funded by the US Department of Energy (DOE) Office of Basic Energy Sciences. Use of the Stanford Synchrotron Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory, is supported by the DOE Office of Science, Office of Basic Energy Sciences under Contract DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research. The authors thank Sebastien Granier for his generous gift of the plasmid encoding BRIL-PTHR, Mark Gladwin for purified Globin-X and Elena G. Kovaleva for H200Q variant of homoprotecatechuate 2,3-dioxygenase crystals. This work was supported by the National Institute of Diabetes and Digestive and kidney Diseases (NIDKK) and the National Institute of General Medical Sciences (NIGMS) of the US National Institutes of Health (NIH) under Award numbers DK102495 (J.P.V), DK011794 (T.J.G), and GM112686 (G.C) and DK102495 (G.C). C.O.B acknowledges support from NIH F31-GM112497. H.P.S and G.C acknowledge support from BioXFEL-STC1231306.

6.0 SERIAL FEMTOSECOND CRYSTALLOGRAPHY USING A X-RAY FREE ELECTRON LASER

In recent years, efforts to design X-ray light sources with unique radiation properties and pulse capabilities has revolutionized biological and physical sciences. The advent of the X-ray free electron laser (XFEL) provides extremely bright pulses on femtosecond timescales, allowing for the probing of biological samples without the conventional radiation damage accrued at synchrotrons. In the remainder of this work, the investigation of methods to improve SFX experiments will be explored. In addition, preliminary results of experimental SFX data collected at LCLS for Pol II transcribing complexes is presented (see **chapter 8**).

6.1.1 The advent of XFEL light sources

X-ray crystallography using synchrotron radiation has been a primary tool for structural studies of large macromolecules. The recent advent of the X-ray Free Electron Laser (XFEL) has advanced the field of x-ray crystallography beyond what is possible from synchrotrons. The Linac Coherent Light Source (LCLS) at Stanford University was the first XFEL to operate in the hard X-ray regime, and came online in 2009 (*189*). XFELs are capable of producing highly brilliant pulses of femtosecond duration and full spatial coherence (*189*). Early thought experiments that simulated the interaction between ultra-short x-ray pulses and biological

molecules suggested that a pulse duration of 10 fs should be short enough to obtain a diffraction pattern while the sample remains intact (*190*). In practice, the high intensity of the X-ray pulse completely destroys the sample upon impact, but the short pulse duration allows for the collection of data before the rearrangement of atoms in the sample can negatively affect the diffraction pattern. This process is referred to as 'diffract-before-destroy', where an interpretable diffraction pattern of nanocrystals, without structural or global radiation damage can be obtained (*169*, *191*).

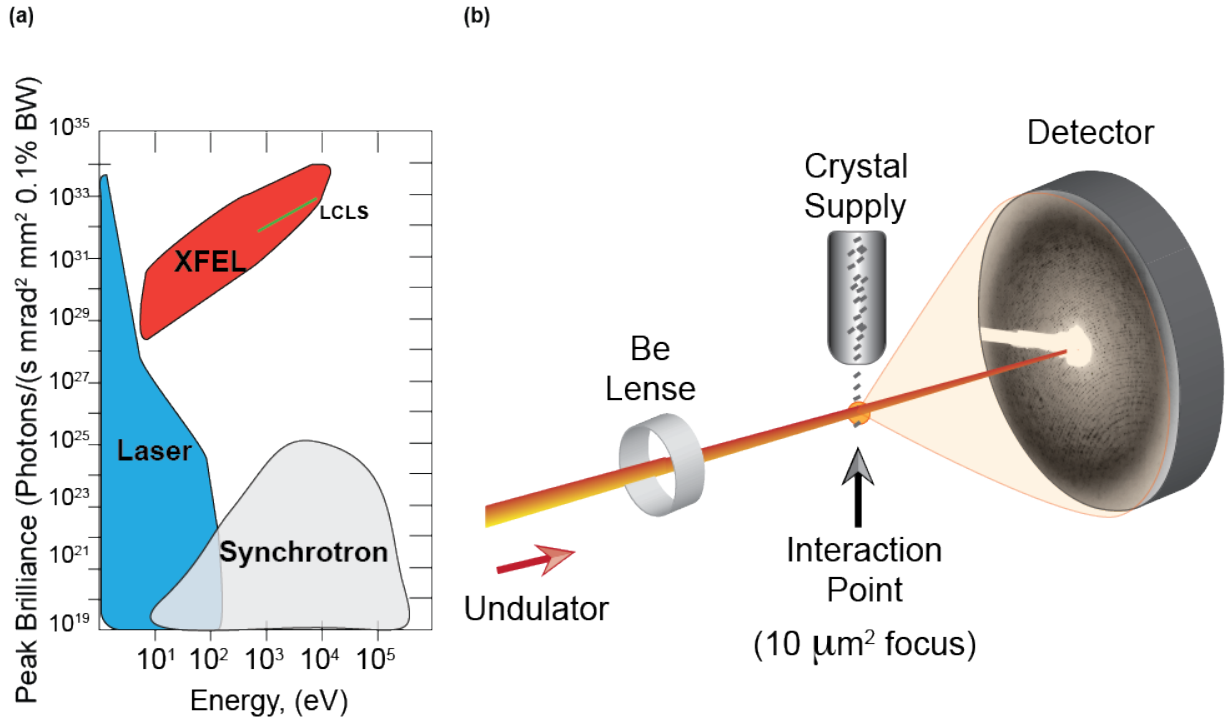

Figure 40. Comparison of peak brilliance and SFX experimental setup

(a) *Peak brightness versus photon energy for conventional lasers and higher harmonic generation sources (Lasers, blue envelope), synchrotrons (grey envelope), and XFEL pulses (red envelope). The green line represents the current spectral range for LCLS beamlines (adapted from Ullrich J, 2012).* ***(b)*** *Schematic of an SFX experimental setup, for the delivery of a continuous stream of nanocrystals across the XFEL source (adapted* (*191*)*).*

These experiments are often carried out in a serial mode using liquid injectors or other devices to deliver uncompromised crystal volumes to the XFEL pulse (Figure 40). Unlike more conventional goniometer-based crystallography experiments carried out at synchrotrons, SFX experiments are contingent upon several thousand indexable diffraction patterns collected from hundreds to thousands of crystals in random orientations to complete a dataset (*169*, *191*, *192*). As a result of the overall setup of the experiment (weak intensities, random orientations and still images) estimation of structure factors is significantly different than rotational X-ray data processing from synchrotrons. Current methods to process SFX data are continuously under development by several groups. One major difference when compared to conventional X-ray experiments is integration with the Monte Carlo method (*192, 193*), where a large quantity of individually integrated hkl intensities of symmetrically unique reflections from different crystals is averaged to correctly estimate structure factor amplitudes. While original data processing methods required hundreds of thousands of diffraction images, improvements in the analysis of sparse datasets (*194-198*), have greatly enhanced the ability to solve structures from XFEL sources.

6.1.2 Delivery of crystalline material

Delivery of nanocrystal slurries into the XFEL path is a dynamic research area with several groups involved in their development (*199-204*). Comparison of the different delivery techniques has been extensively described with great detail elsewhere (*205, 206*). This approach has been highly successful for atomic resolution structure determination yielding several high profile structures (*169, 179, 191, 194, 204, 207*). The most commonly used of these injectors is the gas dynamic virtual nozzle (GDVN) (*208*), that has a volumetric flow rate typically between 1-10

μL/min. Briefly, these injectors provide a continuous small diameter stream of nanocrystal (NC) containing slurry, minimizing background scatter and providing fresh material to each X-ray pulse (delivered at a frequency of 120 Hz at the LCLS). A major drawback of these techniques is that a large percentage of injected material is often wasted passing by between pulses. However, injectors are particularly well suited for time resolved experiments and while most time resolved experiments at XFELs have used lasers to trigger reactions within crystals (*173, 209-211*), a number of mixing injectors are also being developed (*199, 202, 203, 212*). Low flow rate crystal injectors utilizing viscous media such as lipidic cubic phase (LCP), agarose, glycerol and grease (*204, 213-215*) have decreased sample consumption. Moreover, a combination of these slower flow injectors with the intense X-ray micro-beams at micro-focus undulator beam lines and the emergence of fast frame rate area detectors (*216, 217*) have made injector-based diffraction experiments at the synchrotron (*186, 218, 219*) feasible.

Although these developments are promising, a number of issues still complicate the use of injectors, severely compromising experimental success. In particular, injector technologies often require the optimization of carrier solutions to reduce bubbling, crystal clumping and clogging. In many cases crystal slurries are filtered and forced through tiny orifices at high pressures. When these experiments fail to produce diffraction, it is difficult to determine if the problem is related to the inherent quality of the NCs or is a result of solution modifications or the injector protocols employed. In the case of the GDVN injector our previous TEM experiments have shown that the pressures involved in the injection process can be damaging to delicate NCs of multi-protein complexes while other types of NCs, such as of lysozyme, emerge unharmed (*175*).

Notwithstanding the proven success of the injector setups, two possible drawbacks of this methodology commonly arise. First, injector setups often require large amounts of protein to

produce several milliliters of the slurry (containing 20-30% crystals). For some proteins this is a viable option, however for some challenging proteins that require special conditions for expression and purification, this is not practical. Second, there is potential damage that crystals could sustain during the injection process (*175*). Goniometer based fixed-target approaches at the XFEL are advantageous for data collection using delicate crystals, crystals in limited supply, large crystals or for crystal quality screening to prepare for injector-based experiments. The goniometer-based setup (*220, 221*), positions individual crystals into the beam path. However, since the intensity of the beam can create a damaged area around the target spot (20 to 30 µm in diameter depending on the beam diameter), the number of useful diffraction images per crystals is limited. As a consequence, data collection can be time consuming as many crystals need to be positioned and exposed. Efficient grid-based methods have been developed that hold crystals of from 5 to 200 µm in diameter in known locations, in relation to reference marks, for fully automated positioning and data collection (*205, 222*). However, these approaches require careful and often very tedious sample loading steps that could damage delicate crystals. For smaller crystals, areas of sample mounts are often exposed sequentially to search for positions containing crystals (*218, 223*), a time consuming process. Furthermore, crystal non-isomorphism is a serious problem when combining data from multiple crystals.

In the following chapters, research is presented that addresses to concerning areas for both injector-based and fixed target SFX experiments. First, understanding the intrinsic characteristics of NC slurries and their amenability to the injection process is a crucial step to optimize sample preparation and injection protocols for successful SFX experiments and ensure efficient use of the limited beam time available at XFEL sources. To this end, TEM was used to analyze crystalline samples and provide immediate feedback for sample preparation methods to

improve data collection efficiencies. Next, development of an automated approach to efficiently expose low micron sized crystals that conserves crystals and beam time by mapping crystals in random orientations in relation to known reference points on multiple crystal holders (MCHs) is presented. These novel methodologies enabled fast data acquisition and efficient structure solution using minimal sample consumption, while reducing beam time requirements at oversubscribed XFEL and micro-focus synchrotron beam lines.

7.0 ASSESSMENT OF METHODS TO IMPROVE SERIAL FEMTOSECOND CRYSTALLOGRAPHY EXPERIMENTS

7.1 SUMMARY

Serial femtosecond crystallography (SFX) employing high-intensity X-ray free-electron laser (XFEL) sources has enabled structural studies on microcrystalline protein samples. However, the identification and optimization of methods to deliver well diffracting microcrystals with relatively high efficiency remains an experimental challenge. Here, we report improvements in the delivery of microcrystals for both liquid-jet and fixed-target SFX experiments. First, parallel liquid-jet SFX and transmission electron microscopy (TEM) experiments using fragmented microcrystals of wild type (WT) homoprotocatechuate 2,3-dioxygenase (HPCD) and an active site variant (H200Q) revealed traits necessary for efficient sample delivery. Despite identical crystallization conditions and morphology, as well as similar crystal size and density, the indexing efficiency of the diffraction data collected using the H200Q variant sample was over 7-fold higher compared to the diffraction results obtained using the WT sample. TEM analysis revealed an abundance of protein aggregates, crystal conglomerates and a smaller population of highly ordered lattices in the WT sample as compared to the H200Q variant sample. While not reported herein, the 1.75 Å resolution structure of the H200Q variant was determined from ~16 minutes of beam time, demonstrating the utility of TEM analysis in evaluating sample

monodispersity and lattice quality, parameters critical to the efficiency of SFX experiments. In a second set of experiments, we applied novel fixed-target serial femtosecond crystallography methods at the XPP instrument of the Linac Coherent Light Source (LCLS) to reveal the structure of *Escherichia coli* dGTP triphosphohydrolase (dGTPase). We successfully determined the hexameric *apo*-dGTPase structure from 221 still images to 3.2 Å revealing clear Mn^{2+} coordination via the enzyme's HD-motif. These experiments applied UV-fluorescence microscopy, video analysis and highly automated goniometer-based instrumentation to map and rapidly position individual crystals randomly-positioned on fixed target holders. As a result, we achieved the highest indexing-rates observed for a serial femtosecond crystallography (SFX) experiment using randomly positioned micro-crystals.

This work presented in this chapter was adapted with permission from Elsevier Publishing and work pending publication:

Barnes CO*, Kovaleva EG*, Fu X, Stevenson HP, Brewster AS, DePonte DP, Baxter EL, Cohen AE and Calero G. (2016) Assessment of microcrystal quality by transmission electron microscopy for efficient serial femtosecond crystallography. *Arch Biochem Biophys.* **602,** 61-68. doi: 10.1016/j.abb.2016.02.011

Barnes CO, Wu Y, Song J, Lin G, Baxter EL, Brewster AS, Nagarajan V, Holmes A, Soltis M, Ahn J, Cohen AE, Calero G. (2016) Structural basis of enzymatic specificity and inhibition of the dGTP triphosphohydrolase from *Escherichia coli* revealed by XFEL and chemical cross-linking. *(submitted)*

7.2 EXPERIMENTAL METHODS

7.2.1 Quantification of 2,3 – HPCD microcrystal slurries

To quantify the microcrystals, samples were diluted in mother liquor 300 – fold and a 10 µL aliquot was applied to a bright-line hemocytometer (Hausser Scientific). Using a Jan Scientific Jansi UVEX microscope with nominal 5-, 15-, and 40x magnifications the sample chamber grids were visualized using brightfield and UV fluorescence with exposures of 0.1 and 5 seconds, respectively. Images were evaluated using the Jan Scientific CrystalDetect software and nominal 15x magnification images were recorded (Infinity 2-3C Camera, Lumenera Scientific). Manual quantification was achieved using the Vi-CELL XR 2.03 software program (Beckman Coulter). A crystal was defined as having a minimum dimension of 6 pixels (~2.5 µm), with intensity values 3 standard deviations above the background threshold.

7.2.2 Mounting dGTPase crystals on multi-crystal holders (MCHs)

Purification, assembly, and crystallization of wild-type *E. coli* dGTPase was followed as previously described (Stevenson et al., 2016). To mount crystals onto MCHs (see schematic Figure 46), crystals were first cryo-protected with increasing concentrations of mother liquor. Crystal drops were increased to ~10 µL to prevent dehydration during the loading protocol, and MCHs were used to penetrate the drop and extract the randomly oriented crystals. For larger dGTPase macro-crystals (300 x 300 x 200 µm), 8 µL of cryo-solution were pipetted onto the MCH, followed by manual mounting onto the MCH via loop transfer or pipetting. To improve visualization and decrease background diffraction from solvent, excess fluid was carefully

removed from MCHs using filter paper. Optimization of this step is crucial, as wicking of excess solvent may also result in the loss of sample.

7.2.3 Brightfield and UV-microscopy imaging and crystal identification

Brightfield and UV microscopy was employed to identify and locate crystals in relation to the MCH reference points. To this end, MCHs were placed on the stage of the JANSi UVEX UV-microscope, and after focus adjustment, brightfield and UV images covering the entire area of the MCH were acquired with the nominal 5X objective using 0.1 and 1 s exposure times, respectively. After image acquisition, the crystals and MCHs were immediately flash-cooled in a liquid nitrogen bath and transferred into a SSRL cassette. A macro to detect crystals was developed using *ImageJ*, a public domain, Java-based image processing program developed by the NIH (*224*). The macro detects by bright areas that indicate UV-visualized crystals, and outputs crystal profiles and a list specifying crystal size and centroid position (in x-y jpeg pixel coordinates). Optimization of threshold detection, size, circularity and erode parameters are essential for accurate crystal detection. The pixel coordinates of the four MCH reference points were identified in relation to the crystal coordinates through examination of a corresponding brightfield image. A file containing the coordinates of the crystals and the reference points for each MCH was saved in Excel format for input into Blu-ice/DCSS.

7.2.4 Automated data collection using MCHs at the LCLS and dGTPase structure refinement.

Prior to the experiment, a reference file containing crystal coordinates and coordinates of four reference points for each MCH stored in a 96 sample pin storage cassette was read into the *DCS/BLU-ICE* beam line control software (*225*). During the day of the experiment, the SAM mounted each MCH onto the beamline goniometer. The experimenter performed a semi-automated alignment procedure where the MCH is rotated face on to the on-axis microscope and the four reference markers are clicked in a specific order from within a video display of the software interface. Following this procedure the location of the crystal coordinates are displayed over the video image of the mount (Figure 45). The calculated positions are visually inspected and if necessary a graphical interface enables the experimenter to shift the location of the reference points to improve accuracy. A 'start' button is clicked to begin automated data collection. A crystals is translated into the beam position between each X-ray pulse. This process was repeated for each MCH in the cassette.

Diffraction experiments on dGTPase and Pol II complexes were done using 9.5 keV X-ray pulses with 40 fs duration and an 8 μm beam focus at the X-ray interaction point. Diffraction images were recorded on a Rayonix MX325 detector and processed using the *cctbx.xfel* software package (*195*, *226*). Synchrotron based X-ray diffraction experiments of single dGTPase crystals were performed on SSRL beamline BL12-2, APS beamlines 22ID and 23IDD. Data were processed using XDS and SCALA software packages (*227*). Single anomalous diffraction experiments of selenium methionine labelled dGTPase crystals were collected at 12.656 keV with inverse beam every 15° of oscillation data.

Selenium substructure determination was performed using SHELXD, with a resolution cutoff of 4.4 Å corresponding to a CCanom = 0.301. Substructure solutions were utilized in the CRANK2 pipeline (CCP4) resulting in an initial structure of dGTPase which was then manually built in *Coot* and refined in BUSTER. Subsequent dGTPase apo (XFEL) was solved by PHASER (*177*) using the Se-Met structure as a search model. All structures were refined using *Phenix* (*228*) and BUSTER (*105*), followed by several cycles of manual refinement in *Coot* (*109*). All figures were rendered in PyMOL.

7.2.5 Transmission electron microscopy

Microcrystal samples were analyzed by transmission electron microscopy as previously described (*168*). Briefly, 10 μL of sample was applied to glow charged 400 square mesh copper grids with a continuous carbon film (Electron Microscopy Sciences). Following sample incubation on the grid for 30 seconds, excess liquid was removed and the grid stained with 2% (w/v) uranyl acetate. Protein crystallinity of each sample was determined by screening for the presence of lattices. Lattice quality was evaluated by performing a fast Fourier transform of the lattice to obtain the reciprocal lattice reflections or Bragg spots. Bragg spot order of three or above was considered to be of high quality. A single-tile specimen holder was used to deliver samples to a FEI Tecnai T12 electron microscope. The microscope was operated at 120 kV and images were captured with a 2k x 2k Gatan UltraScan 1000 CCD camera.

7.3 RESULTS

7.3.1 Preparation and SFX data collection of HPCD microcrystalline samples

In our first study, we collaborated with the lab of Dr. Aina Cohen at Stanford University to study microcrystalline suspensions of two HPCD stocks, a WT enzyme and an H200Q variant (see **Appendix B**). The single isosteric substitution in the active site (H200Q) has no effect on the crystal morphology (rod-like), spacegroup, unit cell dimensions or diffraction quality as established previously in structural synchrotron-based studies using macrocrystals at cryo-temperatures (*229*). Therefore, these enzyme stocks were used to account for prep-to-prep variability for optimization of the batch crystallization procedures.

Due to rod/needle-like morphology of HPCD crystals and much faster crystal growth in one dimension, fragmentation was required to reduce the length of microcrystals and improve uniformity for injector-based studies. In addition to vortexing (a commonly used fragmentation method described in **section 5.0**) sonication, and centrifugation fragmentation methods were tested (see experimental procedure section). However, because of the limited amount of beam time allocated for this project, diffraction experiments for only two bulk fragmentation methods (centrifugation or vortexing, for H200Q and WT samples respectively), were carried out on the CXI instrument (LCLS) using the gas dynamic virtual nozzle (GDVN) injector (*230*).

Diffraction data for both samples were collected using the same experimental settings, as described in section **Appendix B**. All still diffraction images acquired during beam time were used for indexing and integration with *ccbtx.xfel* (Table 5). Therefore, to enable comparison between samples (Figure 41A,B), the overall “indexing efficiency” is defined here as a

percentage of the total images, collected at the repetition rate of 120 Hz, which could be indexed successfully.

Table 5. Comparison of physical properties and processing characteristics for the HPCD samples used for SFX data collection

	(H200Q HPCD)	(WT HPCD)
Density (crystals/ml)[a]	1.2 x 10^8 (±0.1)	1.53 x 10^8 (±0.07)
Size distribution[b]		
a ~ b (μm)	2.5 (±0.5)	2.6 (±0.6)
c (μm)	11.7 (±5.6)	10.4 (±4.1)
Number of Bragg Spots[c]	4^{th} order	3^{rd} order
Indexing efficiency (%)[d]	12.2	1.7
2^{nd} lattice hits (%)[e]	16.1	8.4

[a] Values for crystal density quantification using nominal 15x magnification (n=12) as described in Experimental Procedure. [b] Size distribution for fragmented microcrystal samples was determined manually from UV imaging data at nominal 15x magnification (n=15). Mean values for short (a, b) and long (c) dimensions for rod-like morphology. [c] Bragg Spot order calculated from FFT transforms of observable lattices (median derived from n=8). [d] The percentage of collected images that contained indexed patterns (single hits and 2^{nd} lattices). [e] Percentage of double hits in the total number of indexed patterns.

During 6.7 min of data collection using the WT enzyme sample (which showed a high tendency of nozzle clogging), a total of 814 diffraction patterns were indexed, corresponding to 1.7% of the total images collected (Table 5). The observed indexing efficiency of 1.7% for the WT enzyme sample compares well with the typical values from previously published SFX studies, where 0.5 – 4% of the still diffraction images acquired at the same repetition rate (120 Hz) were successfully indexed and merged (*169*, *173*, *187*, *191*, *194*, *195*, *231*). In contrast, 12.2% of the total images acquired using the H200Q enzyme sample during 15.6 min of beam time were successfully indexed (Table 1).

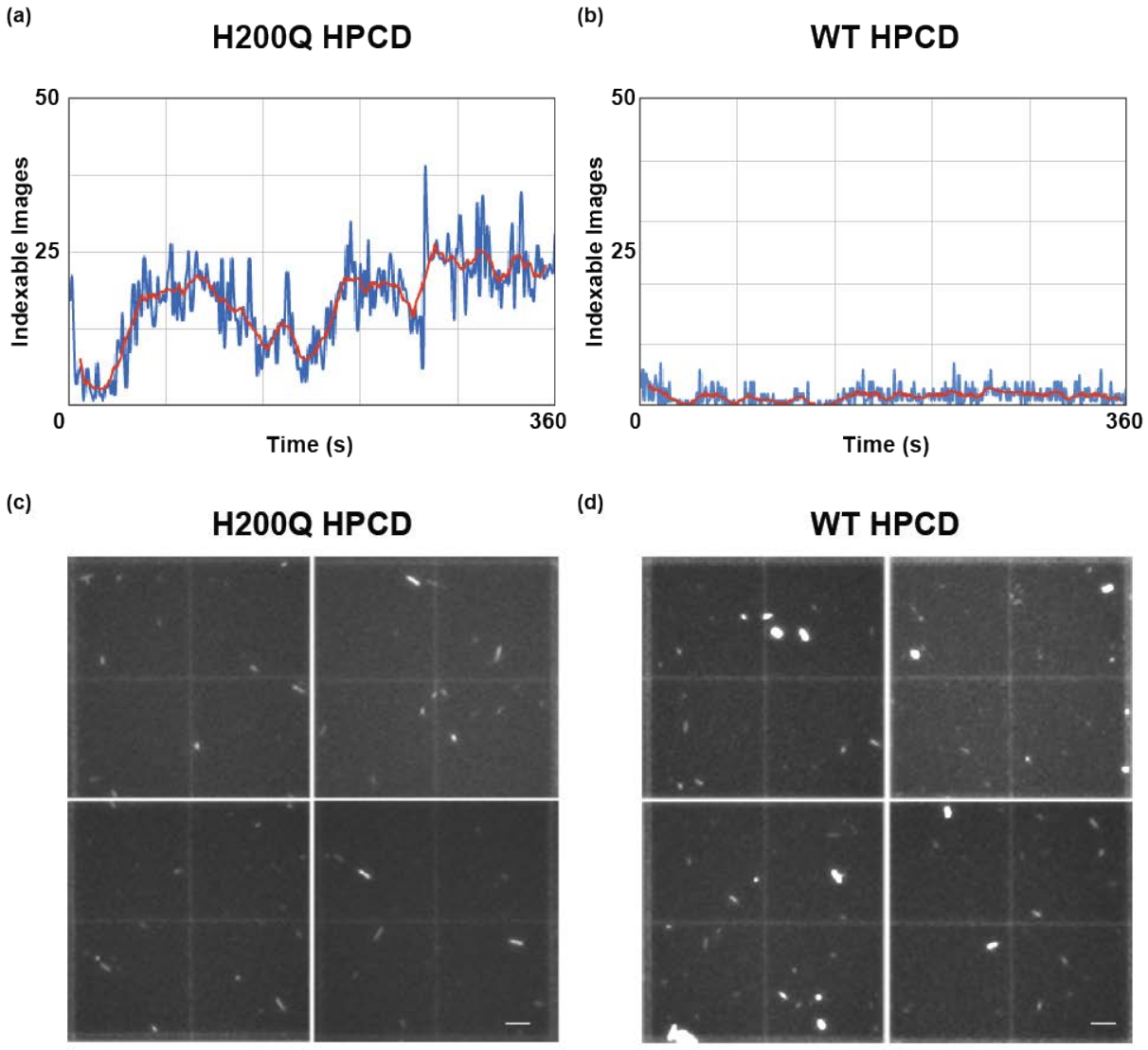

Figure 41. Representative observed indexing rates and crystal density quantification for both H200Q and WT HPCD slurries

*A plot of the number of indexable patterns observed during a representative 6 minutes of data collection using the GDVN injector with the H200Q HPCD microcrystal sample (**a**) and with the WT HPCD microcrystal sample (**b**) illustrates the reduced hit rate observed for the WT sample. The number of indexable hits/second from cctbx.xfel processing are plotted, where the blue trace indicates total number of indexable images per second and the red trace is the averaged indexing rate every 11 sec. Representative hemocytometer quadrants, reconstructed from 4*

*separate images using a 15x objective on a JanScientific Jansi UVEX microscope, illustrate fairly similar crystal densities for H200Q (**c**) and WT (**d**) slurries. Crystal quantification was achieved for both slurries using manual counting with ImageJ software (see experimental section, scale bars – C,D: 50 µm) (n=12).*

7.3.2 Characterizing microcrystalline density and size distribution

We next examined whether differences in the crystal density and size of the microcrystalline slurries could explain the 7.3-fold difference in indexing efficiency of H200Q vs WT, as these are important considerations for sample delivery and sample consumption in liquid injector-based experiments. Previously, crystal densities of ~10^9 crystals/mL have been proposed as the benchmark to attain high hit rates while reducing the occurrence of double hits during GDVN experiments (*191, 210, 232*), although sample densities of 10^{10} to 10^{11} crystals/mL are frequently used as well (*173, 195*). Therefore, in order to quantify accurately the number of crystals for both H200Q and the WT slurries, we applied crystal slurries to a hemocytometer and determined the number of crystals/ml from images obtained during UV-microscopy (Figure 41C,D). Both samples showed similar crystal densities of ~10^8 crystals/ml (Table 5), suggesting that crystal density between slurries was not a contributing factor to the ~7-fold difference in the indexing rate between them. The occurrence of a second lattice was found in 16.1% and 8.4% of indexed images for the H200Q and the WT samples, respectively (Table 5). This data suggests that lower crystal densities may be practical for liquid injector-based experiments.

Furthermore, we examined the crystal size distribution using brightfield and UV microscopy for the WT and the H200Q variant samples (Table 5). In an effort to reduce clogging of the GDVN injector nozzle, samples are typically filtered to limit the maximum

crystal size. If the crystal size distribution of the two samples was significantly different, the different indexing rates observed during data collection could potentially be attributed to increased clogging of the GDVN nozzle, which focuses the crystal slurry to a stream of about 5 μm in diameter. However, our analysis revealed no size difference between the WT and the H200Q variant samples (Table 5). Many of the crystals from both samples were long needles with a significant fraction retaining a length of 10 – 20 μm following the fragmentation and filtration procedures.

7.3.3 Analysis of crystal lattice quality and monodispersity with TEM

The homogeneity and intrinsic order of microcrystals are critical parameters for SFX studies that also cannot be assessed effectively by bright-field microscopy. Similarities between crystal morphology, size, and density raise the possibility that observed differences in the indexing efficiency may be attributed to the quality of the crystal lattice in the two samples. Negative-stain TEM analysis, conducted on the two microcrystalline samples of HPCD, revealed both major differences in the lattice quality and homogeneity (Figure 42). In addition to higher order Bragg spots (> 4^{th} order) observed in H200Q crystals, the EM landscape indicated a highly monodispersed sample (Figure 42A,B). In contrast, WT microcrystals showed a high degree of sample aggregation, protein conglomerates, and a limited number of crystal lattices with > 3^{rd} order Bragg spots (Figure 42C,D). Moreover, delivery of the WT sample was prone to nozzle clogging, which would be consistent with the high degree of protein aggregation observed by TEM and the low indexing rate for this sample (Figure 41B).

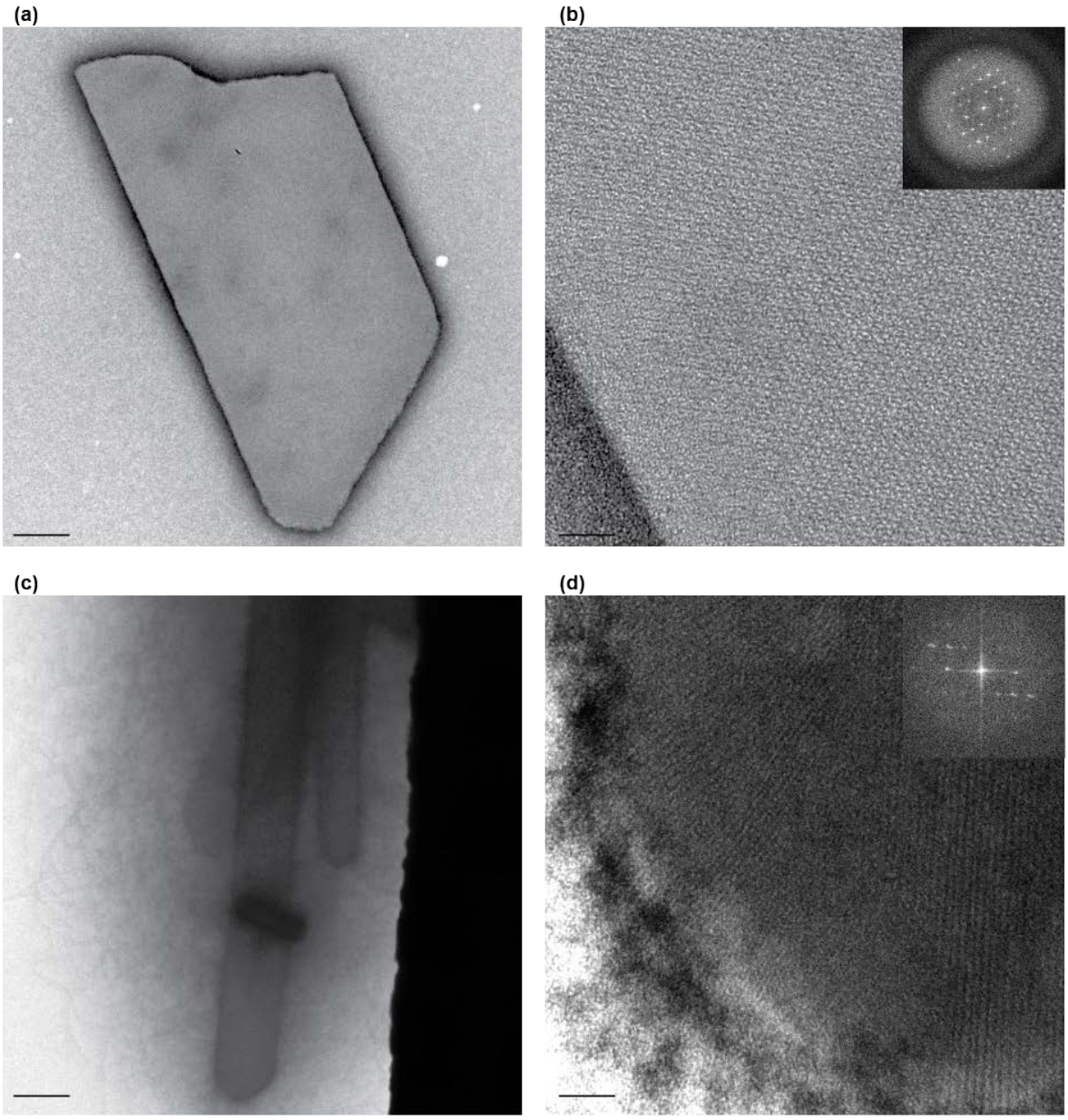

Figure 42. Negative-stain TEM images of HPCD microcrystals

(a,b) *Representative TEM sampling of H200Q crystals, illustrates sample monodispersity (A), with high quality crystal lattices (B) as determined from its FFT profile (inset). On average, 4th order Bragg spots were recorded for H200Q crystals.* ***(c,d)*** *Representative TEM sampling of WT crystals, illustrates sample aggregation or crystal conglomerates (C), typically not seen with brightfield or UV microscopy. Lattice quality was of high quality (D), with an average of 3rd order Bragg spots observed for WT samples. (Scale bars – A,C: 0.5 µm; B,D: 50 nm).*

Since bulk fragmentation of the WT and H200Q microcrystals differed, we performed TEM experiments on the fragmentation protocols to determine its effects on sample monodispersity, lattice quality and crystal aggregation. While we observed no major differences in the lattice quality among HPCD samples prepared by different fragmentation methods, we did observe minor differences in the amount of crystal aggregation. TEM analysis revealed fewer aggregates for crystals fragmented via centrifugation or vortexing (with a glass bead cocktail) when compared to crystals prepared by sonication, irrespective of the characteristics of the starting material prior to fragmentation (Figures 43 and 44). Given that cryo-cooled macrocrystals generated for fragmentation of both H200Q and WT show no difference in diffraction or crystal quality (*229*), the observed variance in the indexing rates between the two otherwise similar samples could stem from the differences in monodispersity and degree of protein or crystal aggregation generated during protein sample preparation or crystallization experiments. The extent and influence of these factors are speculative and will require additional testing to demonstrate a clear significant correlation.

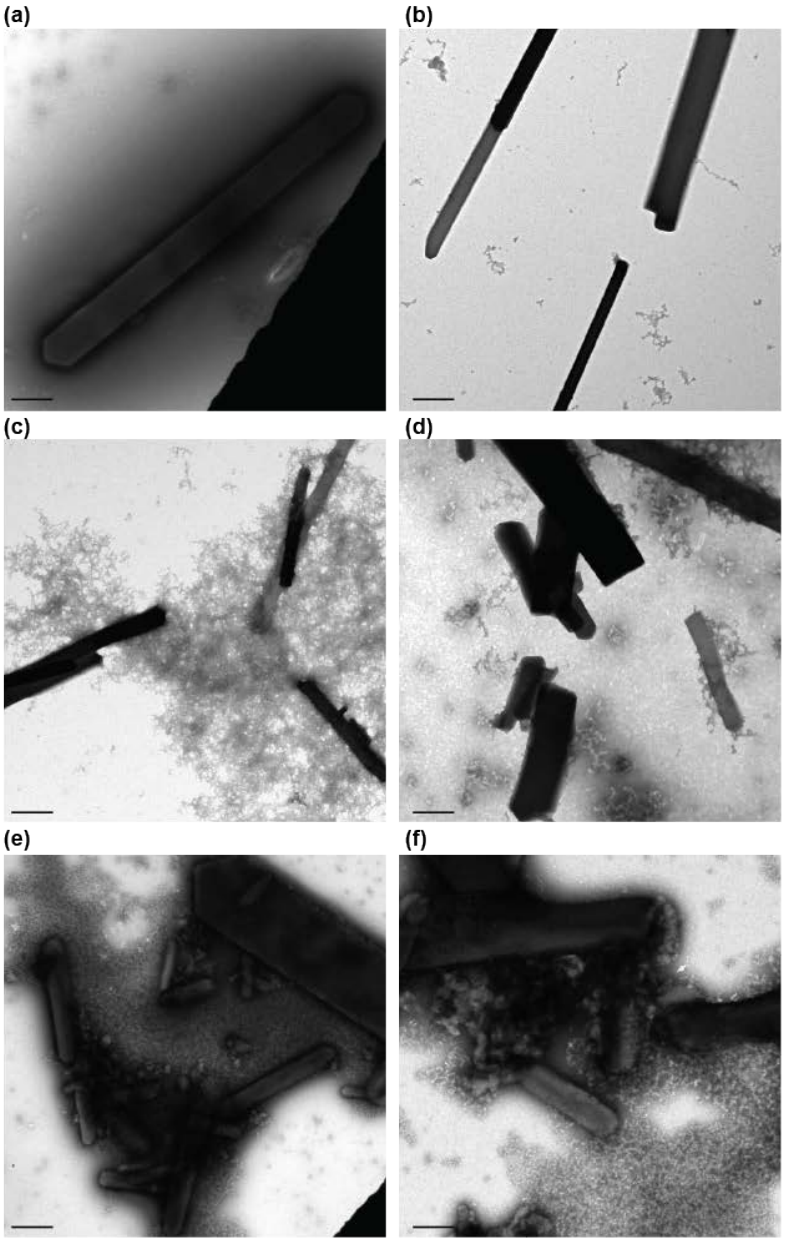

Figure 43. Negative-stain TEM analysis of bulk fragmentation methods for rod-like 2,3 – HPCD microcrystals

(a,b) *TEM images of H200Q (A) or WT (B) crystals after fractionation with centrifugation.* ***(c,d)*** *TEM images of H200Q (C) or WT (D) crystals after fractionation with vortexing.* ***(e,f).*** *TEM images of H200Q (C) or WT (D) crystals after fractionation with sonication. Lattice quality was evaluated when single crystals were observed and with median values of 3rd order Bragg spots (not shown) for all samples across the varying fragmentation protocols. (Scale bars – A-F: 1 µm).*

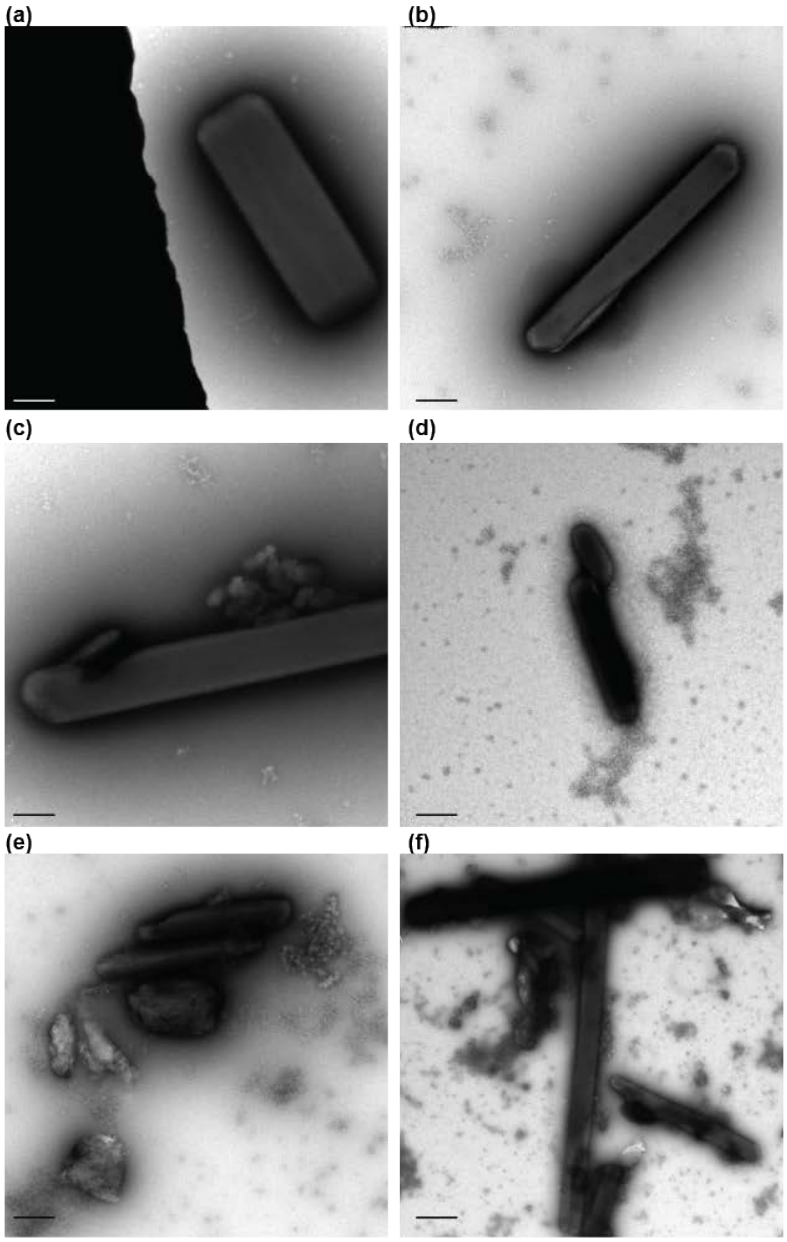

Figure 44. Negative-stain TEM analysis of bulk fragmentation methods for needle-like 2,3 – HPCD microcrystals

(a,b) *TEM images of H200Q (A) or WT (B) crystals after fractionation with centrifugation.* ***(c,d)*** *TEM images of H200Q (C) or WT (D) crystals after fractionation with vortexing.* ***(e,f).*** *TEM images of H200Q (C) or WT (D) crystals after fractionation with sonication. Lattice quality was evaluated when single crystals were observed and with median values of 3rd order Bragg spots (not shown) for all samples across the varying fragmentation protocols. (Scale bars – A-F: 1 µm).*

7.3.4 Mounting, detection, and collection of dGTPase microcrystals on Multi-Crystal Holders (MCHs)

In our second set of experiments using the fixed-target setup at XPP, it was imperative that we develop a device capable of holding 20-100 crystals to eliminate the void time associated with single loop mount/dismount (*222*). We developed a Multi-Crystal Holder (MCH) (Figure 45A,B) capable of holding microcrystals under native- or cryo- crystallization conditions and compatible with with UV microscopy illumination for crystal identification.

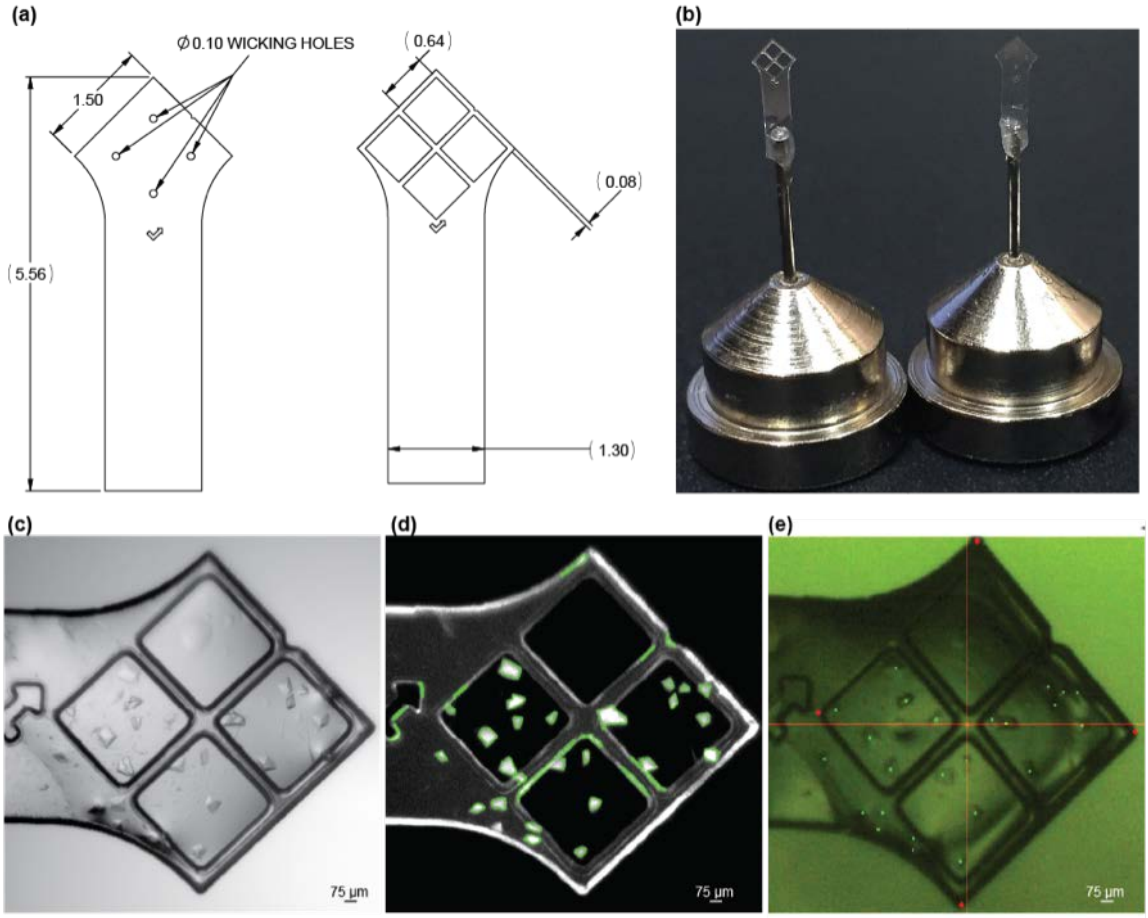

Figure 45. Design and identification of crystals mounted on Multi Crystal Holders

(a) *AutoCAD DXF schematic illustrating the dimensions of MCHs for laser printing onto Mylar sheets.* ***(b)*** *Assembly of MCHs achieved by epoxying laser-printed MCHs onto the pin-base aperture of a Hampton loop.* ***(c)*** *Brightfield*

and, ***d)*** *UV microscopy images of Pol-Spt4/5-DNA mounted crystals. Green outlines indicate the crystal positions determined after particle analysis in ImageJ (ref###).* ***e)*** *BluIce Image of Pol-Spt4/5 crystals mounted at MFX endstation. Coordinates generated in ImageJ (see Table S1) were uploaded, and after identification of reference fiducial marks (red asterisks), beam positions were assigned to each crystal (green boxes).*

Depending on the sample type, the MCHs held 1-2 dozen large size crystals (200-300 µm) to several hundred micro-crystals. The pointed end and large area of the MCH allowed for easy extraction of crystals from the crystallization drops. Brightfield and UV-microscopy imaging was performed at room temperature. Given the high contrast of the UV signal of crystals versus solvent background, detection of crystals on MCHs was easily achieved (Figures 45C and 46B), even for crystals with <10 µm dimensions (Figure 47A-C).

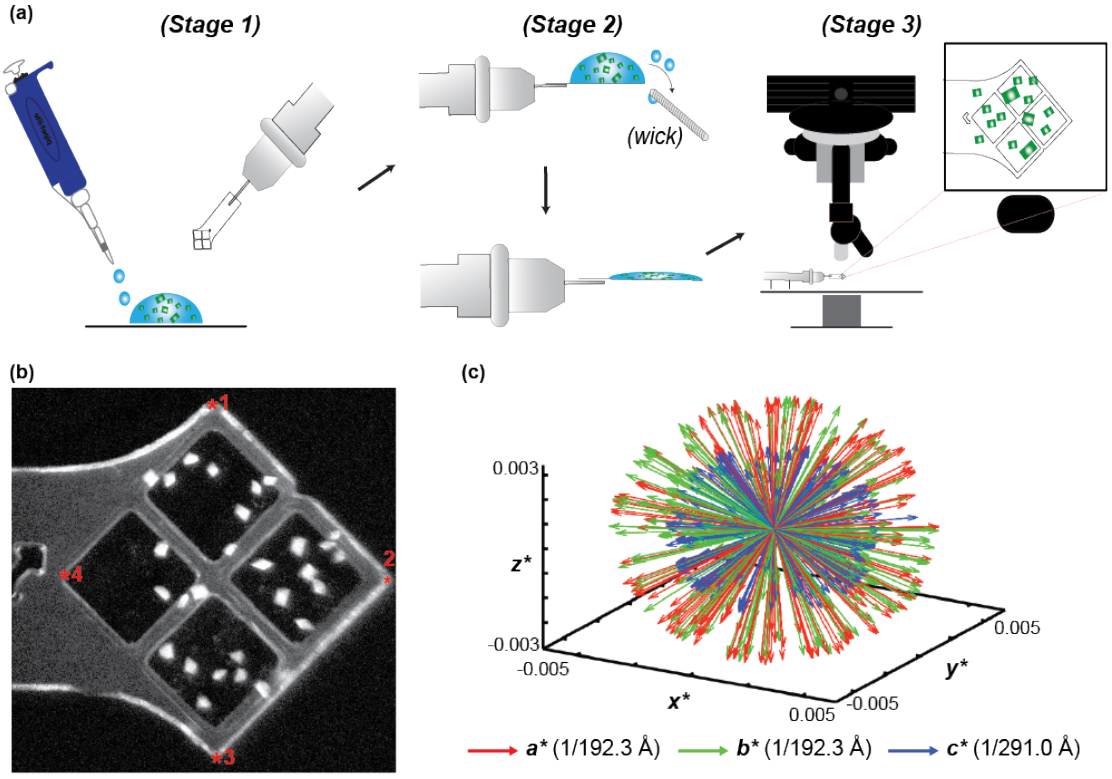

Figure 46. Loading and mapping of crystals on MCHs

(a) *Schematic representation of MCH loading as detailed in the experimental methods section,* ***(b)*** *UV microscopy image of Pol-Spt4/5 mounted crystals. Fiducial marks are illustrated as red asterisk.* ***(c)*** *Reciprocal space*

representation of the basis vectors of 221 indexed dGTPase images demonstrating the lack of preferential alignment when mounting in MCHs.

UV-microscopy images of the MCHs were analyzed to identify bright areas corresponding to the location of individual crystals ((Figure 45C,D). For crystals greater than 100 µm along one dimension (Figure 47D-F), multiple coordinates were recorded, spaced sufficiently to ensure translation into uncompromised regions of the crystal between each X-ray exposure (Figure 47E,F). Algorithms previously incorporated into the Blu-Ice/DCSS experimental control system to position grid-based sample holders and microfluidic traps at LCLS-XPP (*205*, *222*) were adapted to automatically position crystals in random locations based on their relationship to 4 reference coordinates of the MCH for efficient serial diffraction experiments.

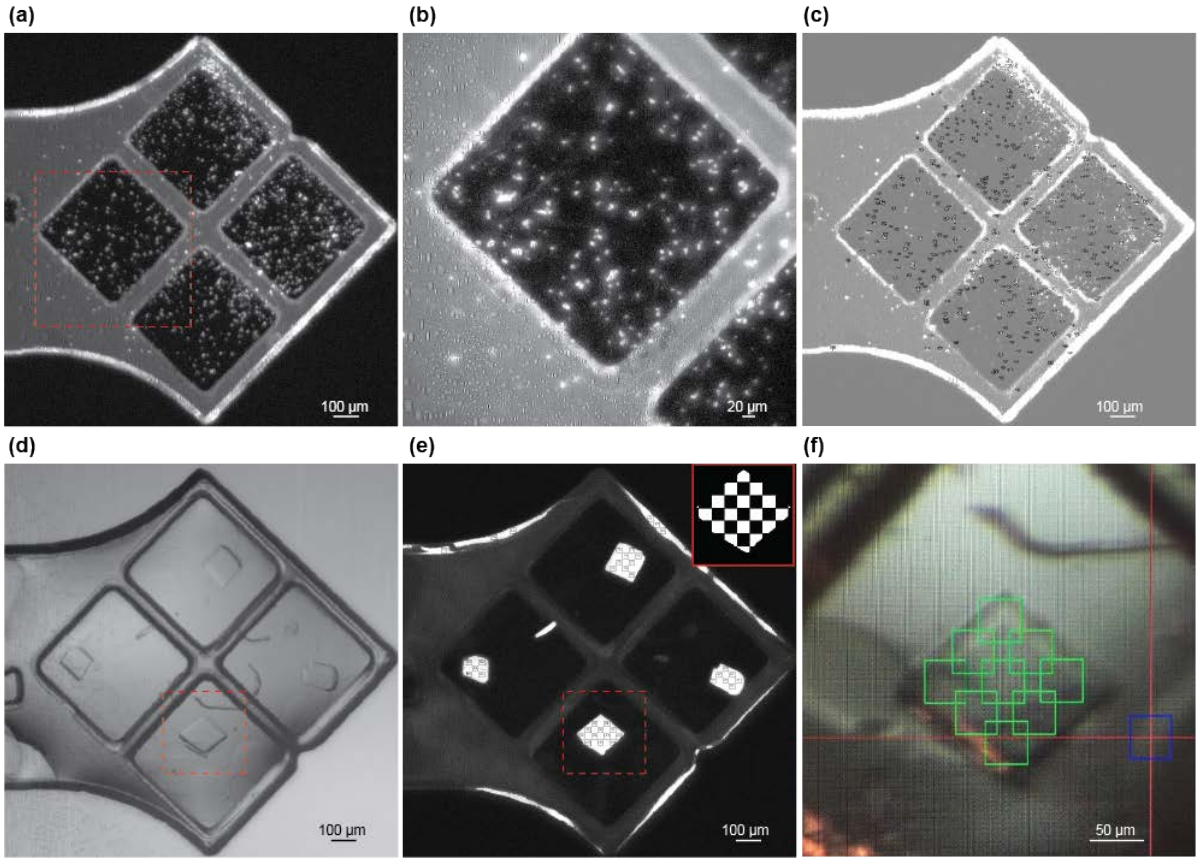

Figure 47. Size distribution of crystals mounted on MCHs

(a-c) *Depiction of greater than 500 E. coli dGTPase microcrystals <10 µm in size, identifiable by UV fluorescence microscopy when mounted on MCHs. and 15x (B) nominal UV magnification images, with ImageJ identification (C) of E. coli dGTPase crystals, measuring <10 µm in size, mounted on MCHs.* ***(d-f)*** *Large crystals of Pol II – TFIIB – DNA illustrating the multi-shot strategy with user specified spacing along the crystals surface.*

This protocol was used successfully with various crystals of complexes mounted on MCH (Figures 46 and 47). Diffraction data collected in an automated fashion using crystals of dGTPase and RNA Polymerase II complexes, resulted in improved intensities of high angle Bragg reflections (Figure 49A), consistent with previous results (*220*). In addition, larger crystals of the Pol II – TFIIB – DNA complex, which utilized a multi-shot strategy with angular offsets for each position (Figure 47D-F), produced indexable still diffraction images. The shots were spaced 50 µm apart, consistent with previously reported spacing protocols (*221*, *233*). Overall, application of MCHs resulted in highly accurate microcrystal hit rates (>80%) for efficient data collection with minimal background scattering preserving weak Bragg reflections.

Moreover, data processing with *cctbx.xfel* produced complete datasets using from 100-1000 crystals, suggesting the micro-crystals oriented randomly on the MCHs. To demonstrate this, we analyzed the reciprocal basis vectors for indexed dGTPase and Pol II crystals that were singly exposed for one MCH (Figure 48) and for the entire dGTPase dataset (Figure 46C) using a method derived from recent observations of a GNNQQNY peptide needle (*234*). Unlike previous goniometer-based XFEL datasets that showed symptoms of preferential alignment of crystals in loops (*221*, *233*), our results revealed a spherical-like projection indicative of randomly oriented crystals. As a result, >90% complete datasets were obtained for all datasets to high-resolution, with only 221 still images being required to solve the structure of the *apo*-dGTPase enzyme.

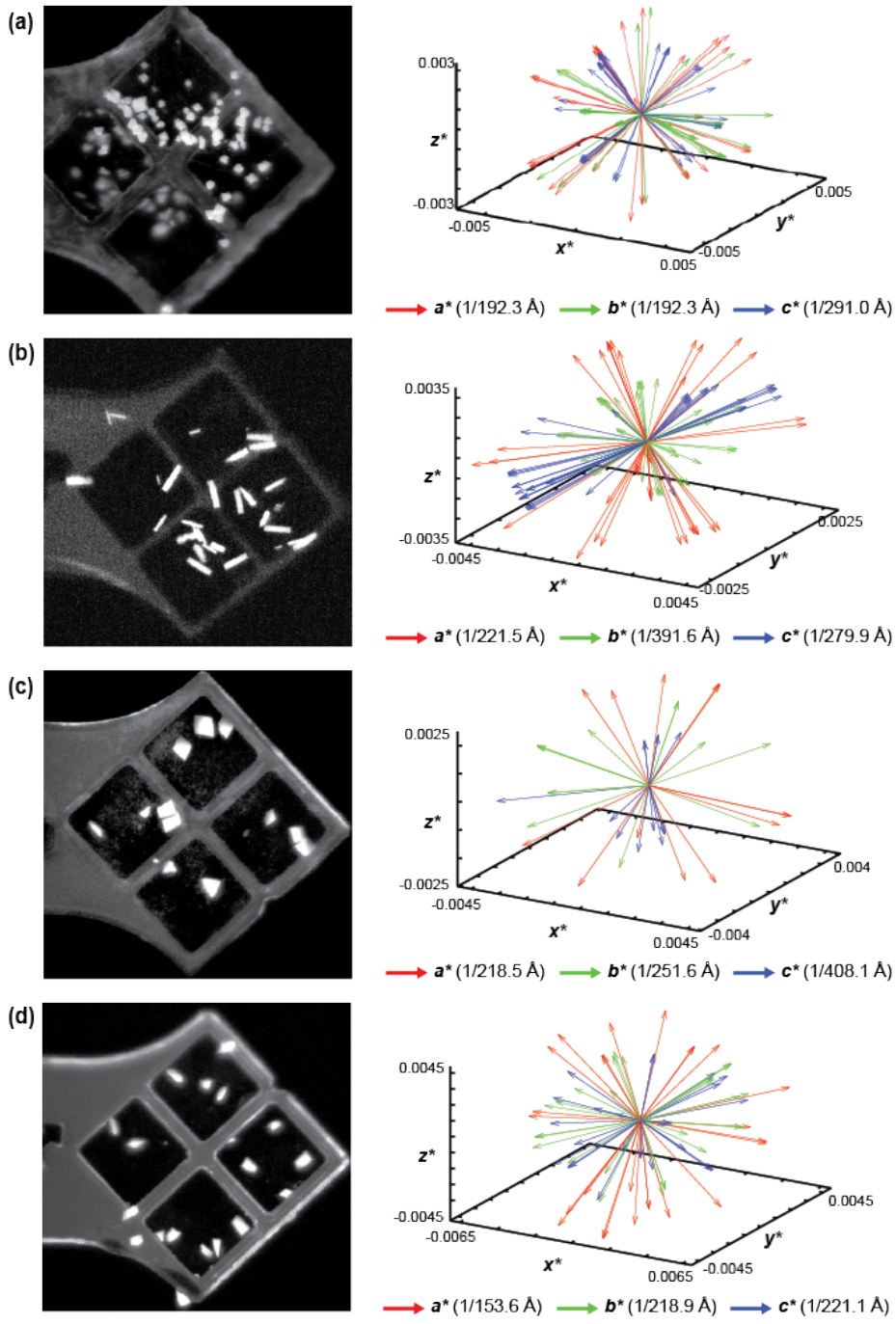

Figure 48. Reciprocal axes analysis of individual grids for tested complexes

*Reciprocal space representation of the basis vectors of indexed crystals that were singularly exposed for **(a)** Ec-dGTPase, **(b)** RNA Polymerase II T834P variant, **c)** RNA Polymerase II – TFIIB – DNA, and **d)** RNA Polymerase II – Spt4/5 – DNA crystals.*

7.3.5 Escherichia coli dGTPase structure and active site metal coordination

The XFEL structure determination of the *E. coli* dGTPase *apo*-enzyme to 3.2 Å was solved from 221 still images captured from ~150 crystals exposed to the XFEL beam (Figure 49), representing the fewest number of still images used to process non-helical data from a goniometer-based XFEL experiment. Initial phases were generated by molecular replacement using a previous selenium-methionine SAD phased structure from our lab as an initial search model. The global fold and unit cell dimensions of our XFEL structure were consistent with previously published dGTPase structures (PDB:ID 4XDS) (*235*), and our synchrotron based *apo*-structures, with an average root mean squared deviation for Cα positions ranging between 0.33 to 0.42 Å when aligned against the refined XFEL structure.

These observed similarities between the Cα positions allowed for the direct comparison of residues involved in active site metal coordination. dGTPase is part of the HD family of nucleases (*236*)with residues H^{69}, H^{117}, D^{118}, and D^{268} participating in metal coordination within the catalytic site. The occupancy for Mn^{2+} divalent cations was set to zero and simulated annealing omit maps were computed for the XFEL data set, to remove model bias for metal ion location based on our starting model. The radiation damage-free XFEL structure showed clear electron density (F_0-F_{calc} map contoured at 3 σ**)** for the divalent Mn^{2+} ion and coordination by the four residues involved in the HD motif (Figure 49B). Most notably, D^{118} was on average 0.5 – 0.7 Å closer to the Mn^{2+} ion across all six monomers when compared to PDB:ID 4XDS (Figure 49C), and our Se-Met *apo*-structure which had an averaged absorbed x-ray dose of 3.4 MGy (*237*). This observed difference, while slight, suggests that at even fairly low x-ray doses, radiation damage is accrued site-specifically around metal centers, possibly through the decarboxylation of the neighboring aspartic acid residue.

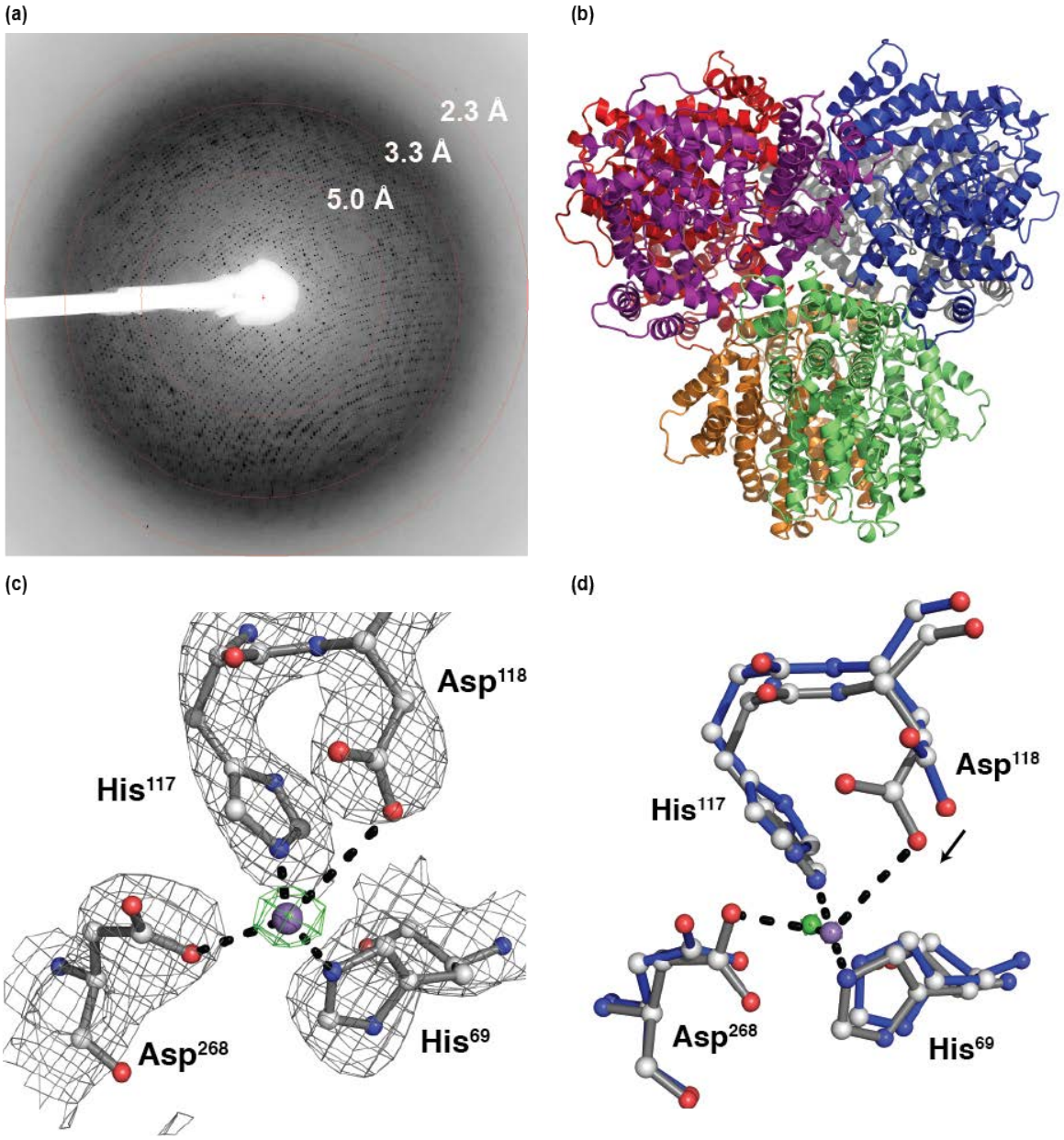

Figure 49. The hexameric dGTPase XFEL crystal structure

(a) *The use of an 11 µm focused beam at XPP extended the resolution of 200 µm crystals from E. coli dGTPase beyond 2.6 Å.* ***(b)*** *Cartoon backbone representation of the hexameric dGTPase XFEL crystal structure solved using 221 still images from randomly oriented crystals,* ***(c)*** *Shown in gray is the $2F_{obs} - F_{calc}$ map contoured at 1.5σ for residues comprising the HD motif and in green the $F_{obs} - F_{calc}$ omit map contoured at 5.0σ for the Mn^{2+} cation. H-bond interactions are indicated with black dashes, all representing distances less than 3.0 Å* ***(d)*** *Differences in Mn^{2+} coordination for XFEL (gray) and synchrotron (blue) structures.*

7.4 DISCUSSION

7.4.1 TEM optimization of samples for injector SFX experiments

Successful SFX experiments require consistent delivery of well diffracting single crystals into the XFEL beam path. For liquid-jet experiments, the inherent diffraction quality of the microcrystals, crystalline monodispersity and the compatibility of the sample to the injection process are all important factors. Our parallel SFX and TEM experiments with HPCD microcrystals illustrate that the use of optimized samples with the GDVN injector, such as the ordered monodispersed H200Q variant microcrystals, will result in efficient SFX experiments. While not reported herein, approximately 16 minutes of protein crystal screening time at CXI (LCLS) resulted in an enzyme structure at 1.75 Å resolution, which is one of the most rapidly collected SFX datasets at atomic resolution using a GDVN injector. In contrast, the presence of protein aggregates or crystal conglomerates can greatly decrease the occurrence of useful diffraction images (even if the crystal lattice is of high quality), requiring much longer data collection time and increased sample volumes to acquire enough data for a complete dataset. Therefore, pre-characterization of samples to identify and optimize conditions that promote monodispersity with highly ordered crystal lattices prior to XFEL beam time would lead to increased efficiency and higher-quality SFX data.

The application of TEM is particularly advantageous in this regard, for systems where crystals are too small (< 5 μm) for X-ray diffraction screening measurements at micro-focused synchrotron beamlines. X-ray powder diffraction (XPD) is a commonly used tool to assess the diffraction capability of microcrystalline preparations, but much like TEM, powder patterns will not exhibit the same resolution to which single crystals diffract at XFEL sources (*231*). Thus,

while this information would be useful where TEM is not accessible, it would not provide details beyond the crystallinity of the sample. The usefulness in TEM compared to other sample preparation tools (I.E. dynamic light scattering, SONICC, UV-fluorescence, XPD, etc.) is that it not only is capable of identifying crystalline material, but also provides the experimenter with additional information about further optimizing or discarding crystals grown under certain conditions (*238*).

In addition to microcrystal quality, our data also raises the possibility that crystal density may play a greater role in data collection efficiency than previously thought. Assessment of our data and previously published results illustrate negative correlation between increasing overall indexing efficiency with decreasing crystal densities (*169, 173, 179, 191, 195, 231, 234*).

Table 6. Comparison of indexing efficiency for similarly solved SFX systems

Sample	Crystal shape/size	Density (crystals/ml)	Number of still images Collected	Indexed	% indexed	Reference
Lysozyme	Box, 1 x 1 x 3 μm^3	1 x 10^{11}	1,471,615	12,247	0.8	Boutet et al., 2012
			1,997,712	10,575	0.5	
PS I	Box, 4 x 4 x 13 μm^3	1 x 10^9	1,850,000	15,445	0.8	(Chapman et al., 2011)
PS II	Box, 2-5 x 15 µm	N/A	1,492,634	4,663	0.3	(Kern et al., 2013)
			586,903	1,848	0.3	
PYP	Needle, 1 x 5 µm	5 x 10^{10}	2,049,294	64,696	3.2	(Tenboer et al., 2014)
			1,024,684	31,255	3.1	
Thermolysin	Box, 2 x 3 x 1 μm^3	2 x 10^{10}	651,793	11,647	1.8	(Hattne et al., 2014)
TbCathepsin B	Needle, ~0.9 x 11 µm	2 x 10^9	3,953,201	178,875	4.5	(Redecke et al., 2013)
2,3 – HPCD	Needle, ~2 x 12 um	1.2 x 10^8	119,040	14,547	12.2	this work
	Needle, ~2 x 11 um	1.5 x 10^8	47,882	814	1.7	this work

Furthermore, when considering the monodispersed H200Q sample (as visualized by TEM) which had lower crystal densities than those routinely used for GDVN based experiments, we still observed ~20% double lattice hits for indexed images. This raises the possibility that lower crystal densities may be advantageous and practical for liquid-jet experiments. However,

since TEM experiments to analyze monodispersity and overall quality of crystal lattices were not performed for previously published samples, we cannot rule out that these factors played a role in their overall indexing hit rates. Yet taken together, these data suggest that monodisperse samples analyzed by TEM with crystal densities of 10^8 crystals/mL may improve efficiency and lessen sample consumption, which is applicable to challenging protein targets (I.E. membrane proteins, multi-protein complexes, etc.).

7.4.2 Improved hit-rate efficiencies using MCHs at fixed-target SFX setups

Notwithstanding the proven success of the injector setups, two possible drawbacks of this methodology commonly arise. First, injector setups often require large amounts of protein to produce several milliliters of the slurry, and second, there is potential damage that crystals could sustain during the injection process either from incompatibility with carrier solutions or the shear forces involved in the process (*175*). Goniometer based fixed-target approaches at the XFEL are advantageous for data collection using delicate crystals, crystals in limited supply, large crystals or for crystal quality screening to prepare for injector-based experiments. Our data presents a highly efficient fixed target data collection methodology that provides greater than 90% crystal hit rates for serial crystallography using randomly oriented micrometer-sized crystals. While crystals in MCHs were mapped prior to the experiment using UV-microscopy at the home laboratory, future experiments will incorporate a UV-imaging capability on the beamline goniometer for on-the-fly crystal identification and mapping. This will improve the precision of crystal positioning, and providing a straightforward means to fully automate these fixed target SFX experiments using a variety of MCH form factors. As UV-microscopy imaging is broadly applicable to identify a wide-range of macromolecular crystals, the general crystallographic

community may easily adopt this fully automated approach for multi-crystal experiments at both synchrotron and XFEL sources. As crystals in random locations may be automatically positioned, this method supports methods combining *in situ* growth and data collection.

7.5 ACKNOWLEDGEMENTS

This work was supported by SLAC and Laboratory Directed Research Development grant 0006-15 (E.G.K. and A.E.C.). Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, and Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. We thank Michael M. Mbughuni for protein purification, and staff at LCLS-CXI for assistance with data collection. This work was supported by the National Institute of General Medical Sciences (NIGMS) of the US National Institutes of Health (NIH) under Award numbers GM112686 (G.C), GM112497 (C.O.B), and GM102520 (A. S. B.). H.P.S and G.C acknowledge support from BioXFEL-STC1231306.

8.0 STRUCTURAL BASIS OF THE GAIN OF FUNCTION POL II T834P VARIANT REVEALED BY A XFEL

8.1 SUMMARY

Goniometer-based serial femtosecond crystallography (SFX) employing high-intensity X-ray free-electron laser (XFEL) sources has enabled structural studies of crystals in the 20-100 μm range revealing highly significant protein targets. Here we present preliminary results for the use of an automated, highly efficient data collection protocol for fixed-target SFX experiments, enabling the collection of several RNA Polymerase II transcribing complexes, including the gain of function (GOF) allele T834P Pol II, with resolution to ~3.3 Å. Using both synchrotron and XFEL sources for data collection on Pol II microcrystals, three independent datasets show that T834P allows capture of Pol II with a kinked bridge helix (BH), induced by the proline substitution, and traps the TL in closed conformation with a matched NTP substrate at the highest resolution observed so far. Our structure reveals stronger TL-BH interactions, a rearrangement of Met^{1079} in the TL hydrophobic pocket, and other changes in our closed TL relative to published structures. These structures, in conjunction with future molecular dynamics (MD) and biochemical studies may reshape our basic understanding of TL function and reinforce the need for new structures at higher resolution.

8.2 BACKGROUND

Since the pioneering works of the Kornberg lab in the early century, our wealth of knowledge on eukaryotic RNA Polymerases has exponentially increased from structures revealing the underlying mechanism of transcription. The transient interplay between nucleic acids, Pol II and the transcription factors that regulate transcription activity have been illuminated in recent years from both X-ray crystallography (*80*, *81, 239*) and single particle cryo-EM models (*90-93*). As the molecular mechanisms of transcription initiation and elongation are emerging, questions still remain regarding the allosteric role these factors play in the nucleotide addition cycle (NAC) and translocation. Specifically, it remains to be seen how the basal transcription machinery might affect trigger loop (TL) activity during early transcribing events.

As previously discussed in the above chapters, the TL is a highly conserved loop comprising Rpb1 residues 1078-1097, and plays a fundamental role in the NAC (*56*, *57*, *59*, *240*). Evidence from single molecule studies on Pol II indicates that alteration of TL residues influences Pol II translocation properties, consistent with models presented in **Chapter 3** (*241*). In particular, TL opening is proposed to be a critical step during Pol II translocation across the DNA template, and as we previously showed, over 40 potential H-bonds across the NAS must be disrupted for this to occur. Our previous work also revealed that interactions with the non-transcribing DNA strand (NTS) and the C-terminal proximal region of the TL (termed the U-loop) may play a role in TL stability in an open state. In addition, it is possible that the numerous transcription factors which engage the NTS (ie. TFIIB or TFIIF in the early initiation complexes), may also influence TL conformational states, as evident by the partially folded TL in the presence of TFIIF in a Pol II transcribing complex (*241*).

Moreover, the importance of interactions between Pol II proximal domains and TL conformational states has emerged in recent years. Kinking of the highly conserved bridge helix (BH), which exists in all multi-subunit Pols, has been proposed to support translocation. In the archaea *Methanocaldococcus jannaschii* (Mja) RNAP, substitutions of Met^{808} or Ser^{824} for a proline at two hinge-proximal residues resulted in gain of function (GOF) phenotypes, suggesting that the introduction of a proline residue kinks the helix, resulting in increased translocation or catalysis (*52*, *242*, *243*). A recent study in *S. cerevisiae* with BH mutants analogous to Mja RNAP variants (Rpb1 Met^{818} and Thr^{834}) showed gain of function phenotypes with similar transcriptional properties as Mja RNAP (*198*). The higher rate of *in vitro* elongation suggests that these single substitutions promote a TL conformation similar to the closed state. Thus, structurally understanding how distinct TL conformations or interactions are linked to different functions in transcription, possibly by incorporating signals from the rest of Pol II or Pol II bound factors (*198*, *244*), is important in revealing TL dynamics.

In this preliminary work, we sought to identify the mechanism which governs gain of function transcriptional activity of the Pol II T834P variant in the presence of a complete and short nucleic acid scaffold. Owing to the the importance of the metal ion in nucleotide catalysis (*46*, *49*, *240*), Pol II T834P crystals were a suitable candidate for data collection at the Linac Coherent Light Source (LCLS) to garner a radiation damage-free structure. Herein, we describe the utilization of MCHs, to solve the high-resolution crystal structure of the radiation-damage free Pol II – T834P variant in the presence of ATP to 3.3 Å, revealing the closed TL. Furthermore, coupling changes in the closed TL state with atomic-resolution data collected at the synchrotron for Pol II-T834P crystals in the closed and open TL states, we provide detailed mechanistic insights into TL function and dynamics. Our structures reveal stronger TL-BH

interactions, a rearrangement of Met^{1079} in the TL hydrophobic pocket, and other changes in our closed TL relative to published structures. In addition, we demonstrate improvements to data resolution and electron density map quality for radiation-sensitive crystals of RNA Polymerase II (Pol II) complexes collected at the LCLS, consistent with our previously observed results.

8.3 EXPERIMENTAL METHODS

8.3.1 Protein purification, assembly, and crystallization

The single point mutation T834P was introduced into *rpb1* using site-directed mutagenesis as previously described (*56*, *198*). Wild type Pol II and T834P variant were purified as described above (section **3.1.1**). The transcription factor IIB and elongation factors Spt4/5 were purified as previously described (*220*, *238*). To assemble Pol II transcribing complexes, synthetic oligonucleotides shown in Table 1 (i+5 NAS for TFIIB and Spt4/5), containing a single stretch of non-complementary bases and a 9-mer RNA (5' – AUCAGAGGA – 3' for Spt4/5) or 5-mer RNA (5' – GAGGA – 3') TFIIB construct) were annealed above their melting temperatures and allowed to cool slowly to 4º C. Synthetic oligonucleotides were similarly annealed for Pol II – T834P complexes with the following sequences, TS: 5'-CTACCGATAAGCAGACGTCCCTCTCGATG, NTS: 5'-CTGCTTATCGGTAG, and a 10-mer RNA with 3´- end dG (5' – AUCGAGAGG/3´dG/ for T834P variant). The resulting nucleic acid scaffolds (NAS) were mixed with Pol II (3:1 molar ratio) and excess scaffold was removed using size exclusion chromatography (Superdex200, GE LifeSciences) against Buffer A (25 mM Hepes pH 7.5, 100 mM NaCl, 5 mM DTT, 0.5 mM EDTA, 10 µM $ZnCl_2$). A 2.5-molar excess of

TFIIB or Spt4/5 was mixed with the Pol II-TIC in the presence of 1M NaCl, and dialyzed overnight at 4º C against Buffer A. Removal of excess TFIIB or Spt4/5 was carried out using a second size exclusion step (Superdex200, GE LifeSciences) against Buffer A. Purified assemblies were concentrated to 6-8 mg/mL and crystallized in the following conditions:

Table 7. Pol II complex crystallization conditions for SFX experiments

Complex	Crystallization condition
Pol-II-T834P-s10	25-27 % Tacismate pH 7.0, 0.1M Hepes pH 7.0, 1% glycerol, 5mM $MgCl_2$
Pol-II-TFIIB-DNA	28-30 % Tacismate pH 7.0, 0.1M Hepes pH 7.0, 1% glycerol
Pol-II-Spt4/5-DNA	10-12% Peg 4000, 75 mM $MgCl_2$, 0.1 M MES, pH 6.0

^ All conditions include 10 mM DTT or 5mM TCEP, and where cryo-protected with increasing concentrations of mother liquor and glycerol.

For the addition of matched nucleotide into Pol II – T834P – s10 crystals, prior to cryo-cooling 15 mM ATP was added to the cryo-buffer to initiate TL folding. Crystals were soaked in the presence of ATP for ~15 minutes before flash-freezing in liquid N_2.

8.3.2 XFEL data collection and processing

Crystals of Pol II complexes were mounted onto MCHs as described in section **7.2.2**. Diffraction experiments were carried out at XPP and MX beamlines (LCLS) on Pol II complexes using 9.5 keV X-ray pulses with 40 fs duration and an 8 µm beam focus at the X-ray interaction point. Diffraction images were recorded on a Rayonix MX325 detector and processed using the *cctbx.xfel* software package (*195, 197*). To improve the accuracy of our data, a grid search of spot-finding parameters on an image-to-image basis was performed to maximize the success of indexing and integration (*245*). Diffraction images were then divided into individual clusters

based on their crystal symmetry and unit cell parameters using hierarchical clustering with default (5%) and a very narrow cutoff of 0.5% tolerance for unit cell parameters (*246*, *247*). Using the largest cluster, we employed post-refinement as implemented in the program PRIME (*248*) to generate a merged diffraction data set. Synchrotron based X-ray diffraction experiments of single Pol II - T834P – s10 crystals were performed on SSRL beamline BL12-2 and APS beamline 23IDB using an X-ray energy of 12.665 keV. Data were processed using XDS and SCALA software packages (*227*).

8.3.3 Structure refinement and analysis

The structures were solved by molecular replacement (MR) using 12-subunit Pol II PDB:ID 5C4A (*241*) for Pol II-T834P or Pol II-Spt4/5 complexes, or PDB:ID 4BBS (*81*) for Pol II-TFIIB complex in Phaser (*104*). MR models were initially refined in CNS (*108*) using rigid body and B-group refinement, followed by atomic displacement parameter (ADP) restraints. Pol II-T834P models were iteratively built with B-factor sharpening in COOT (*109, 110*) and further refined in the programs Buster (*105*), and Phenix (*228*) using TLS and ADP restraints. All figures were rendered using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

8.4 PRELIMINARY RESULTS AND DISCUSSION

8.4.1 Pol II crystals benefit from higher XFEL intensities at fixed goniometer setups

As previously described, we successfully used TEM to identify (*168*), assess (*175*, *238*), and optimize (*249*) crystallographic studies of nanocrystals. Therefore, initially, we tested the ability of Pol II complex nanocrystals to diffract using liquid-jet delivery methods as described in chapter **7.** Utilizing nanocrystals of the Pol II – TFIIB – DNA complex and the GDVN delivery method (*208*, *230*), crystals were diffracted at the CXI endstation of LCLS. While diffraction was observed (Figure 50A,B), resolution was typically limited to 4.0 Å or less. This was surprising given that slightly larger sized crystals diffracted to 3.7 Å with long exposure times (40 seconds) and the use of a 10 X 10 μm beam size using a synchrotron light source. These results suggested that 1) diffraction was dose dependent and 2) the possibility of accumulating crystal damage during jet injection due to the high solvent content of the crystals, as we previously observed (*175*).

However, given the ability to generate larger crystals (Figure 38B), we next tested whether improvements in resolution could be achieved from cryo-cooled crystals mounted on MCHs at fixed-goniometer XFEL endstations. Remarkably crystals of the Pol II – TFIIB – DNA complex diffracted beyond 3.2 Å (Figure 50C,D), while crystals of the Pol II – Spt4/5 – DNA complex also saw improvements in the observation of high-angle Bragg reflections. On average, improvements of 0.3 – 0.6 Å resolution were observed, consistent with previously observed work (*220*, *233*). In addition, crystals of the Pol II T834P variant soaked with a matched nucleotide were also collected, with consistent diffraction to 3.3 Å or better.

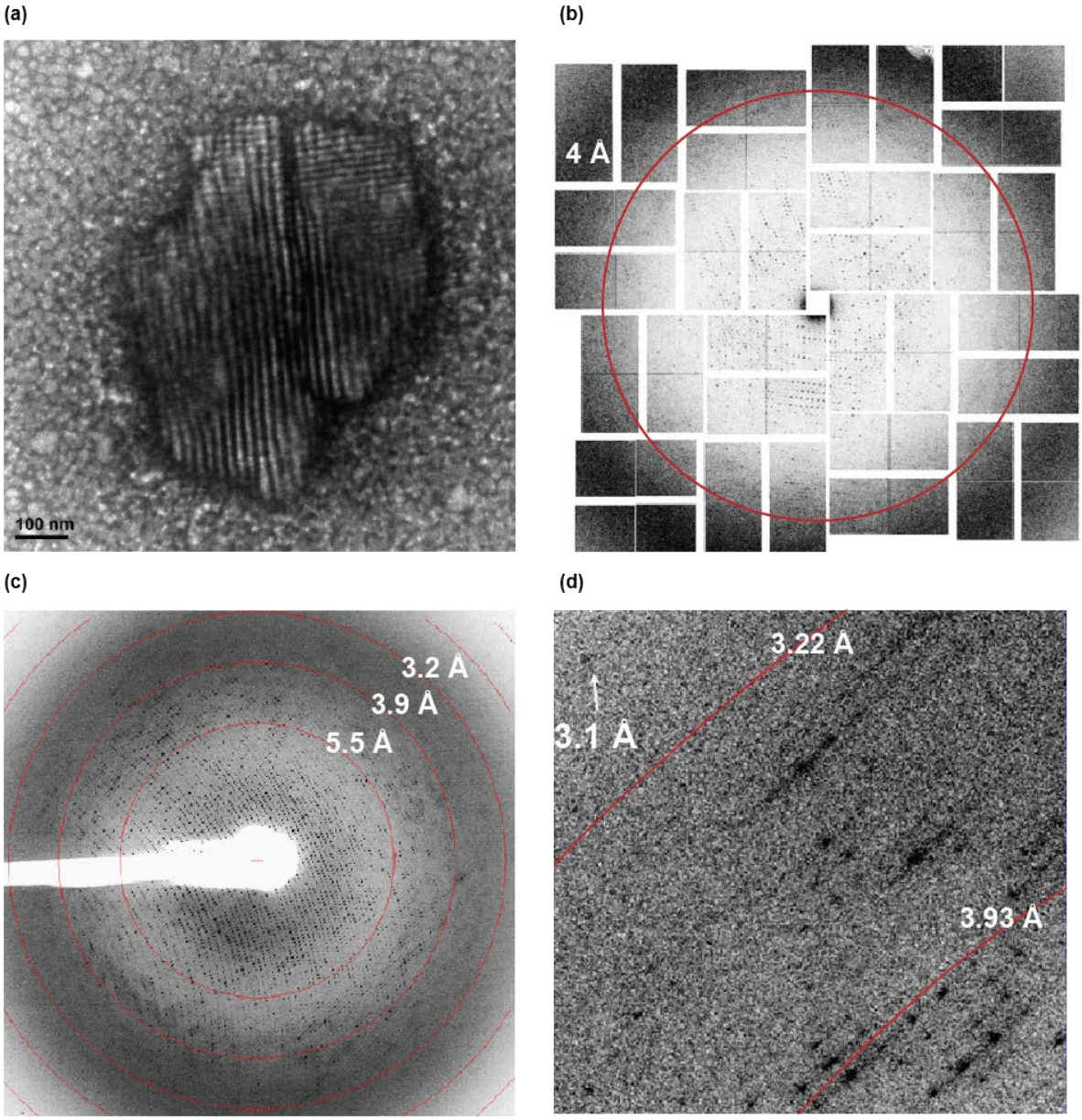

Figure 50. Diffraction of Pol II - TFIIB - DNA crystals at XFEL sources

(a) *Fragmented crystals of Pol II – TFIIB – DNA complex were prepared and injected using the GDVN.* ***(b)*** *diffraction of nano-crystals on a CSPAD detector at CXI.* ***(c)*** *Representative diffraction of Pol II – TFIIB – DNA crystals at XPP endstation, with higher resolution* ***(d)*** *being clearly observed beyond 3.2 Å.*

8.4.2 Analysis of data collection methodologies, merging statistics, and initial map features of SFX generated data

As described in section **8.3.3**, *cctbx.xfel* was used to index, integrate and merge our XFEL data (*195*, *197*). To determine the efficiency of data collection using MCHs, we defined hit rate as the number of images with observable diffraction (using a cutoff of 100 Bragg reflections during image-by-image grid search in *IOTA* (*245*) against the total number of previously unfiltered images for each dataset. Our observed hit rates of ~90% are among the highest hit-rates ever reported for single-shot XFEL diffraction experiments (Table 8).

Table 8. Data collection and processing statistics for Pol II SFX experiment

	Pol II – T834P – s10 + ATP (LCLS - XPP)	*Pol II – Spt4/5 DNA (TB_{25}) (LCLS - XPP)*	*Pol II – Spt4/5 – DNA (TB_{18}) (LCLS – XPP)*	*Pol II – TFIIB – DNA (LCLS – XPP)*
Experimental Parameters				
Average Crystal Size (*a x b*, μm)	80 x 80	60 x 100	40 x 70	100 x 100
Total Number of Images	1085	613	807	574
Total Hits (% of total images)	984 (90.6)	523 (85.3)	643 (79.7)	507 (88.3)
*Indexed/Merged (% of total images)	883 (81.4)	421 (68.7)	594 (73.6)	394 (68.6)
Space Group	$C222_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Resolution (Å)[b]	35-3.3	35-3.85	35-3.5	35-3.4
*Completeness (%)	97.9 (91.3)	98.3 (97.5)	99.8 (97.2)	91.6 (90.7)
Redundancy	7.3 (3.3)	7.8 (5.4)	9.3 (6.1)	5.5 (3.3)
$\langle I/\sigma I \rangle$	6.1 (0.8)	5.7 (0.4)	4.8 (0.3)	4.6 (1.0)
R_{split} (%)	37.3 (85.2)	32.9 (88)	34.9 (98)	56 (106)

*Merging with default unit cell tolerance of 5%

Using the largest cluster of indexed images, we employed post-refinement as implemented in *PRIME* (*248*) to estimate structure factor amplitudes. When using a default setting of 5% unit-cell tolerance and inclusion of negative intensities (*197*, *250*), 10-20% of images were rejected, resulting in >90% complete datasets for all complexes collected (Table 8). However, due to the non-isomorphism of the crystal samples, standard deviations along the *a*, *b*, & *c* axes were extremely high (10-15 Å on average, as high as 28 Å for Spt4/5 and TFIIB datasets along the longest axis), negatively affecting map features of mobile elements (Figure 51). As a result, subsequent hierarchal clustering of images was restricted to a unit cell tolerance of 0.5% to improve crystal isomorphism, but at the detriment of dataset completeness. Such narrow cutoffs rejected on average 50-60% of images from the original merged datasets, with only the Pol II – T834P dataset showing >90% completeness better than 4Å resolution (Table 9). Yet, this approach was critical to reveal the complete TL in the Pol II – T834P transcribing complex as shown in Figure 51C,F.

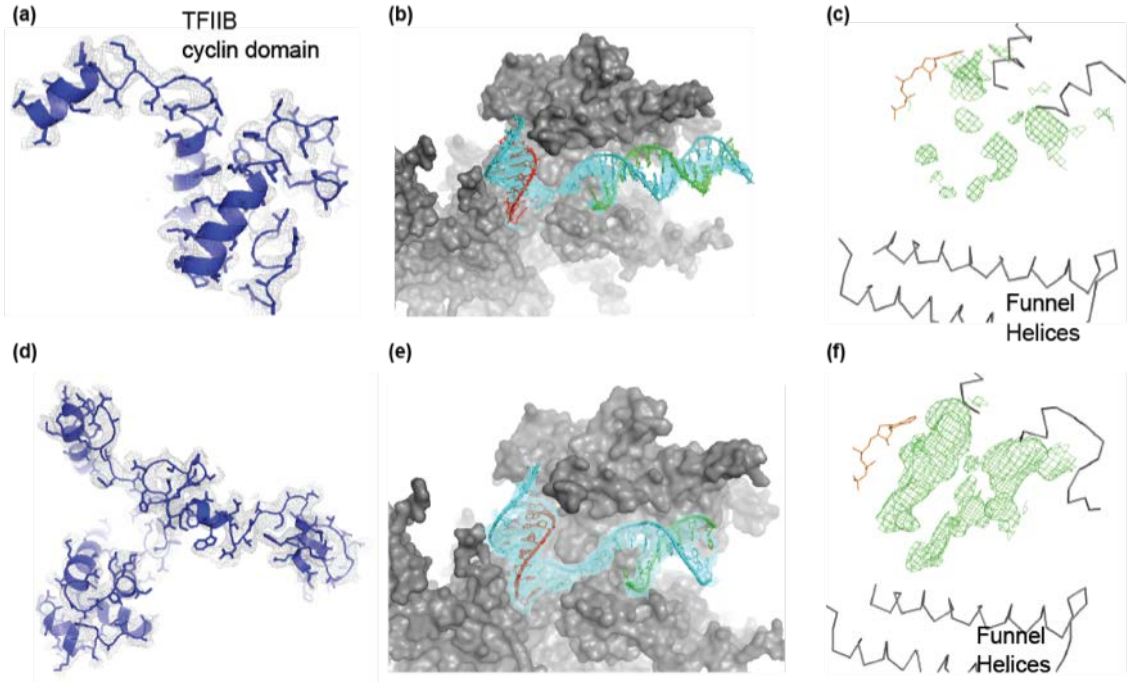

Figure 51. Map quality improvements as a function of unit cell tolerance

Improvements in electron density were observed for all complexes tested when unit cell tolerance was restricted from the default 5% (a-c) to a user-defined setting of 0.5% (d-f). ***(a,d)*** *$2F_{obs} - F_{calc}$ initially refined map contoured at 1σ for TFIIB cyclin domain,* ***(b,e))*** *$2F_{obs} - F_{calc}$ map contoured at 1σ for nucleic acid scaffold in Pol II – Spt4/5 crystals, and* ***(c,f))*** *$F_{obs} - F_{calc}$ difference electron density contoured at 2σ for the trigger loop (green mesh) in Pol II – T834P crystals. The matched ATP nucleotide is shown in orange.*

Table 9. Pol II T834P data reduction and refinement statistics

	Pol II – T834P – s10 +ATP (LCLS - MX)	*Pol II – T834P – s10 +ATP (SSRL - 12-2)*	*Pol II – T834P – s10 (no ATP soak) (APS - 23IDB)*
Data Collection[a]			
PDB:ID	**n/a**	**n/a**	**n/a**
Space Group	$C222_1$	$C222_1$	$C222_1$
Unit cell (Å)	220.2, 392.5, 280.9	221.1, 391.6, 279.9	221.0, 393.9, 283.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.304	0.9798	0.9798
Resolution (Å)[b]	35-3.3	40-3.67	40-3.45
Unique Reflections	188,170	133,292	162,143
Completeness (%)	97.9 (91.3)	100 (100)	99.8 (99.7)
Redundancy	7.3 (3.3)	9.3 (9.1)	10.2 (8.8)
$CC_{1/2}$ (%)	82.8 (25.5)	99.0 (28.9)	99.8 (60.1)
$\langle I/\sigma I \rangle$	6.1 (0.8)	4.7 (0.9)	5.5 (0.7)
Mosaicity (°)		0.2	0.35
R_{merge} (%)	64.7 (84.5)	29.9 (248)	33.7 (302)
R_{pim} (%)	37.3 (85.2)[c]	10.2 (86.5)	12.2 (94.1)
Data Processing Program	cctbx.xfel[d]	XDS/Scala	XDS/Scala
Refinement			
No. Atoms	31,984	32,035	31,461
R_{cryst}/R_{free} (%)	29.5/31.1	18.9/22.5	22.6/25.5
Refinement Program	Phenix/CNS	Buster/CNS	Buster/CNS

[a]Numbers in parentheses correspond to the highest resolution shell

[b]Resolution limits were extended to include weak intensity data (Karplus and Diederichs, 2012). Using the traditional criterion of $I/\sigma I > 2.0$, the resolution limit is 3.72 Å, 3.95 Å and 3.82 Å, respectively.

[c]Value represents R_{split} (%) for data collected at XFEL source, not R_{pim}.

8.4.3 Structure of Pol II - T834P variant: Bridge helix and trigger loop conformation

Prior to XFEL data collection, single crystal diffraction experiments of a Pol II – T834P transcribing complex soaked with a matched nucleotide were collected at beamlines GM/CA (APS) and 12-2 (SSRL) to 4.0 Å and 3.65 Å, respectively. These structures were solved by molecular replacement using PDB:ID 5C4A. To prevent model bias during initial refinement, Rpb1 residues 670-740 (Funnel helices), 815-850 (Bridge Helix, BH), and 1072-1100 (Trigger Loop, TL) were removed from the molecular replacement model. The refined 3.65 Å structure was subsequently used for molecular replacement of the XFEL data, followed by simulated annealing using DEN refinement (*251*) in CNS. The models were further refined using the programs *Buster* (*105*), *Phenix* (*228*) and manual building with B-factor sharpening in COOT (*109*, *110*).

The BH and TL domains were iteratively built into the F_{obs}-F_{calc} electron density map using the characteristic features of bulky aromatic side-chains as register (Figure 52A,B). Initial analysis of the BH reveals a kinked conformation, not only in the crystals with the matched ATP present (Figure 52A), but also for a recently collected single crystal dataset of T834P without ATP soaking (Figure 52B). Overlay between these data, with BH domains from previous TL on-state (PDB:ID 2E2H, (*49*)) or TL off-state (PDB:ID 5C4J, (*241*)) revealed that the T834P substitution promotes a BH conformation similar to TL-on state, even when a matched nucleotide is not present (Figure 52C,D). This suggests that the T834P variant does not undergo a conformational change (kink to straight) in the BH when transitioning from the TL on/off states.

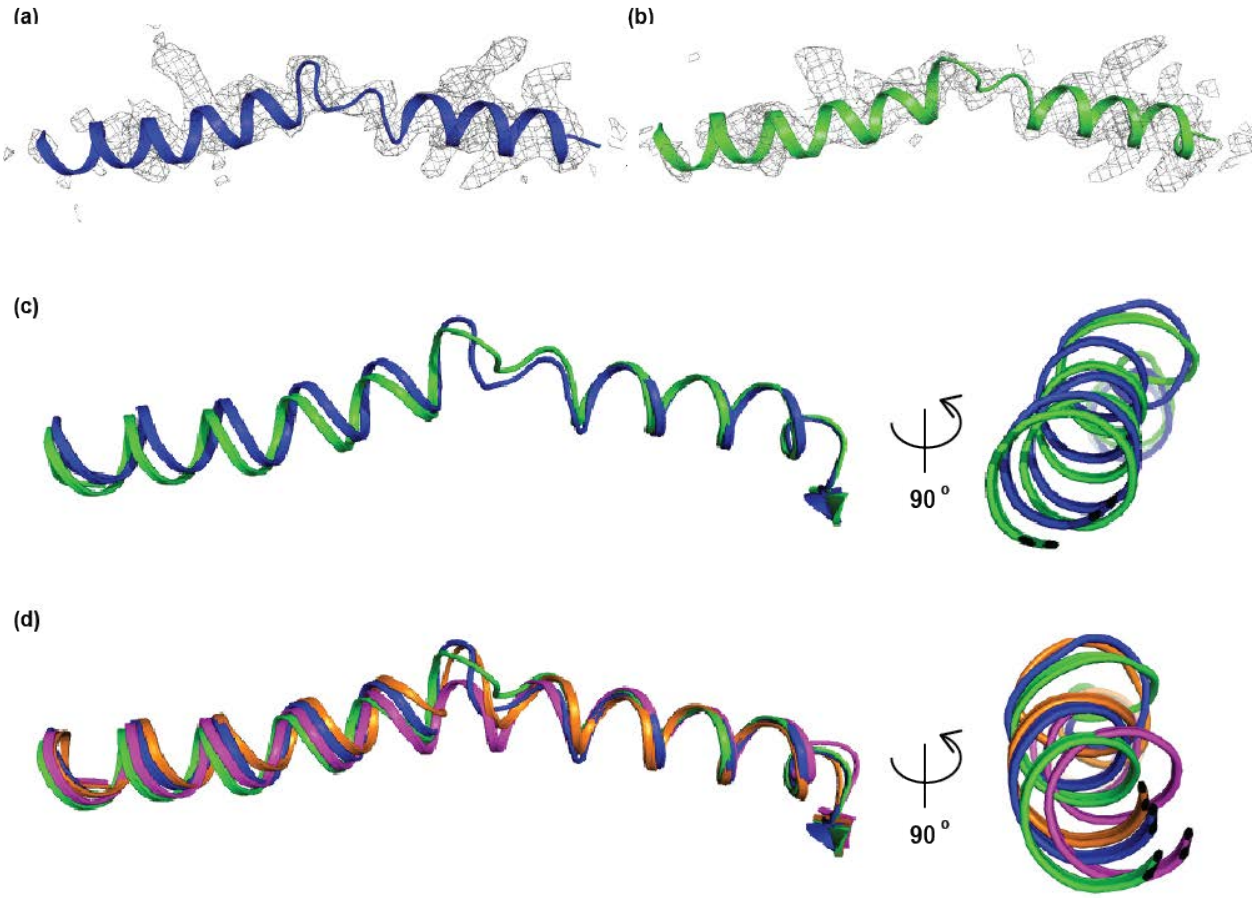

Figure 52. Comparison of Bridge Helix conformations

(a) *$F_{obs} - F_{calc}$ electron density map contoured at 3σ of Rpb1 residues 815-845 for T834P crystals soaked with ATP (blue),* ***(b)*** *and without ATP (green).* ***(c)*** *Overlay between the two datasets reveal strikingly similar conformations, with only slight deviation in Cα positions.* ***(d)*** *In comparison, overlay with PDB:ID 2E2H (orange) and 5C4J (magenta), shows that the Pol II T834P variant does not undergo a conformational change between TL-on (orange) to off (magenta) states as does the wild type enzyme.*

8.4.4 Met^{1079} remains buried during on-state transition

As mentioned above, the $F_{obs} - F_{calc}$ electron density map showed characteristic features of the TL on-state when a matched ATP nucleotide was present in the active center (Figure 51F). The

final refined $2F_{obs} - F_{calc}$ electron density contoured at 1.5σ revealed in great detail the TL (Figure 53A).

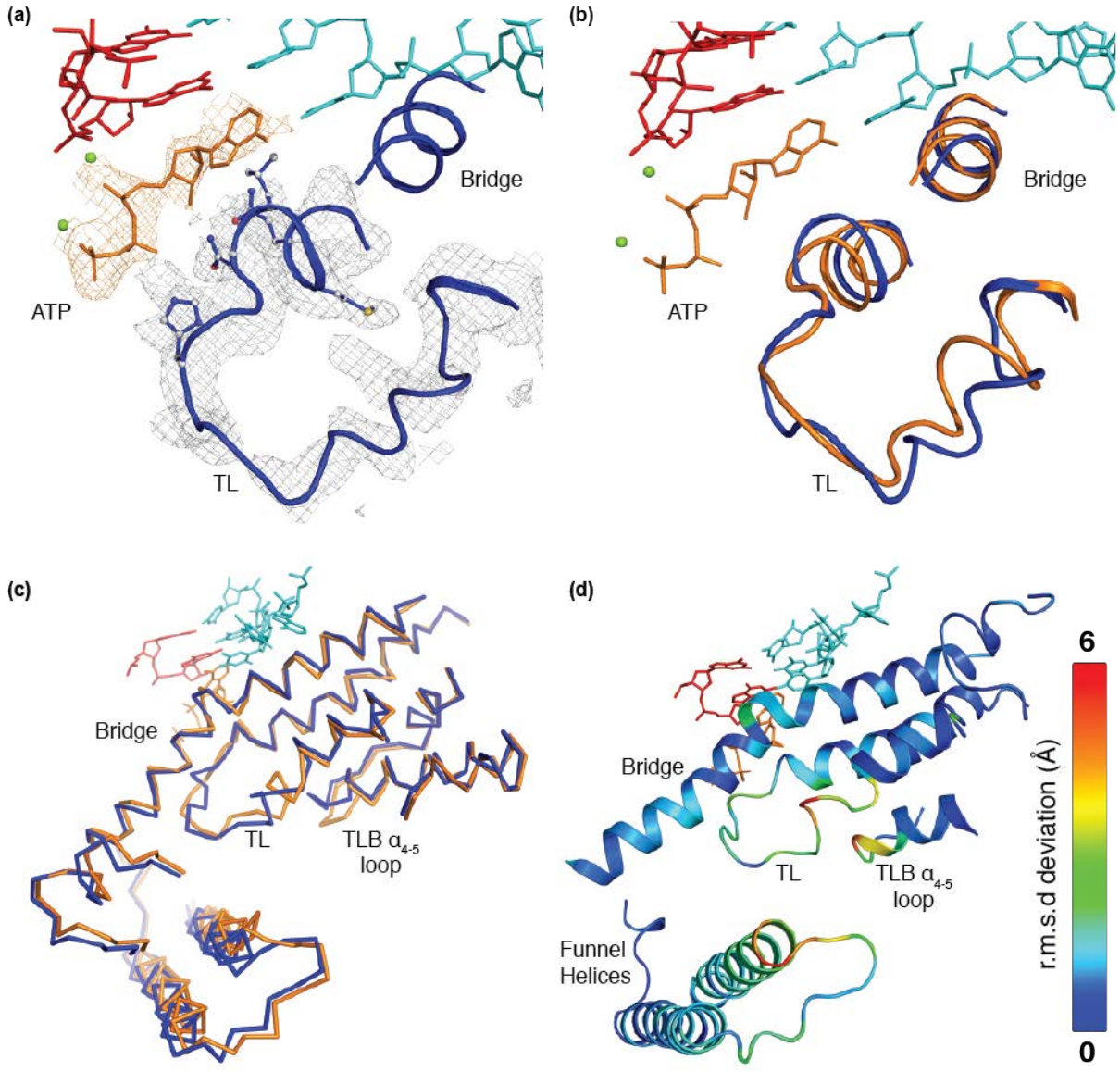

Figure 53. TL density and comparison with previous on-state structure

(a) *The final refined $2F_{obs} - F_{calc}$ electron density map of TL residues (grey mesh) captured in the on-state engaging the matched ATP nucleotide (orange) in the A site with divalent Mg cations. Template strand DNA (cyan) and neighboring RNA nucleotides (red) are also shown.* ***(b)*** *Overlay of our structure (blue) with PDB:ID 2E2H (orange), illustrating the similarities in TL and BH residues.* ***(c)*** *Ribbon representation of overlay active site*

residues showing that local similarities do not extend to surrounding Pol II domains. ***(d)*** *VMD analysis of Cα deviations of our structure aligned against 2E2H.*

In addition to continuous density for TL residues, we also observed strong density for ATP occupying the A site (Figure 53A,B) as previously reported (*49, 60*). Overlay between our structure with PDB:ID 2E2H (*49*) showed a similar architecture of the BH and TL domains (Figure 53B). However, while the global fold of the active site was similar average root mean squared deviations for Cα positions ranged from 0.44 – 5.8 Å (Figure 53C,D), with the funnel helices showing some of the largest deviations (average deviation of 2.4 Å).

Strikingly, the observation of Rpb1 residues 1355-1364 in a closed conformation was surprising given that this loop acts as a lid to a highly conserved hydrophobic pocket at the core of the TLB (*241*). Our structures revealed the presence of Met^{1079} situated in the hydrophobic pocket at the core of the TLB (Figure 54A), a feature that underscores previous observations that the burying of Met^{1079} inside this pocket is indicative of the TL-off state (*49*, *55*, *57*, *59*, *241*). Genetic interactions and gene expression profiling of substitutions of Rpb1 Ala^{1076}, Gly^{1097} and Leu^{1101} with residues that disrupt the Rpb1 Met^{1079} hydrophobic pocket (*127*), support a model where the integrity of the pocket is critical for stabilization of the off state in eukaryotic and archaeal polymerases. Such substitutions are highly related to those that hamper off state conformations by destabilization of the C-terminal TL hinge region; therefore, these substitutions are similarly predicted to alter translocation rate and catalysis (Kaplan et al., 2012; Kaplan et al., 2008; Kireeva et al., 2012; Malagon et al., 2006; Wang et al., 2006). So while pocket integrity is necessary for the stabilization of TL off-state, our data suggest that disruption of the TLB pocket is sufficient, but not necessary for TL-on state conformational transitions.

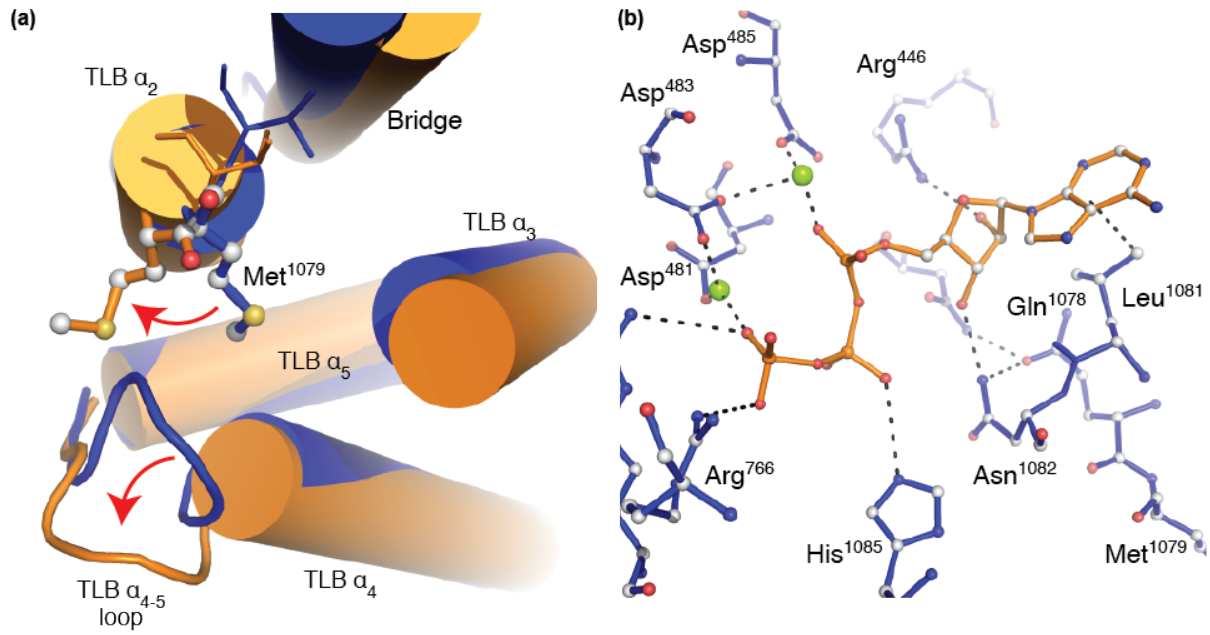

Figure 54. Position of Met[1079] and interactions with ATP

(a) *Overlay between our structure (blue) and 2E2H shows that our structure was able to capture the TL on-state without disruption of the TLB core. Indeed, Met[1079] stays buried in the hydrophobic pocket, stacking against a loop connecting TLB helices 4&5.* ***(b)*** *Interactions between residues of the TL and proximal Rpb1 domains with ATP in the A site.*

While surprising, this observation is not without precedence. Previous structures (PDB:ID 4A3F) for a partially folded TL in the presence of a matched nucleotide in the A site showed a nucleotide interacting region (NIR) similar to our structure, without disruption of the TLB by Met[1079] (*60*). Moreover, the structure of an elongating *T. thermophiles* bacterial enzyme in the presence of a matched nucleotide also showed a homologous leucine buried in the hydrophobic pocket of the TLB, despite an on-state TL conformation (*47*). It is possible that we trapped an intermediate in which the observed H-bonding between Asp[1359] of the TLB loop with Lys[1093] of the TL C-terminus maintains pocket integrity, which is supported by TL simulation models in which the Lys-Asp interaction is considered a transition intermediate directly before

the on-state is reached (*252*). While speculative, further simulations and biochemical data should be considered to gain a better understanding of TL dynamics.

8.4.5 Functional observations of TL nucleotide interacting region (NIR)

The trigger loop engages in a network of interactions with ATP in the A site along with proximal Rpb1 domains lining the active site (Figure 54B). Trigger loop residues in the nucleotide interacting region (NIR) including His^{1085}, Asn^{1082}, Leu^{1081}, and Gln^{1078} all directly engage the ATP (Figure 54B). Coordination of the ATP phosphate region by His^{1085} engaging the β-phosphate, while Aspartic resdiues 481, 483, and 485 of Rpb1 coordinate the two metal ions. Interestingly, while the first metal was discovered to engage the α-phosphate, metal two was found in the corresponding C-metal site coordinating the γ-phosphate, as previously observed (*25*, *49*). While the roles of NIR residues are unchanged in the moieties they recognize compared to our understanding of active site architecture, our structure revealed a different H-bonding network compared to previous structures of a fully folded TL (Figure 55) (*49*).

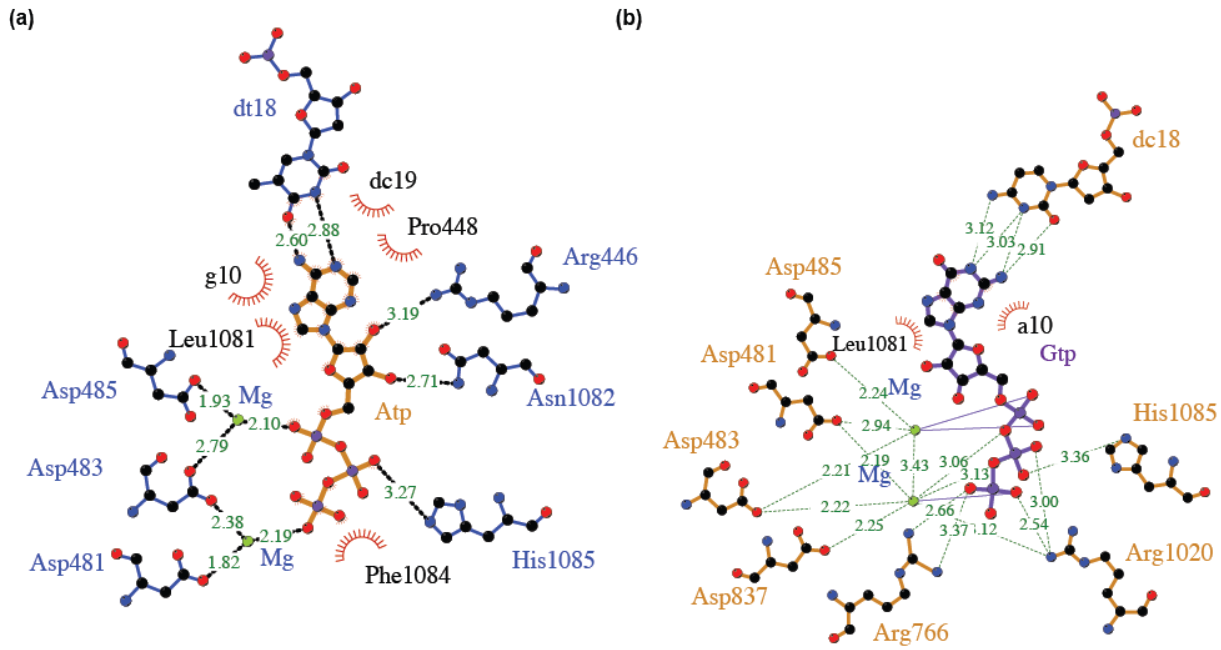

Figure 55. LigPlot analysis of NTP interactions in our structure versus 2E2H

(a) *LigPlot analysis of interactions between Pol II with ATP in our structure, and* ***(b)*** *Pol II with GTP in PDB:ID 2E2H.*

In addition to the metal positions, Asn[1082] and Arg[446] were shown to engage the ribose 3´-OH and 2´-OH groups respectively, a feature not seen in 2E2H. Engagement by the 2´-OH by Arg[446], along with Asn[479] (not seen in our structure), was observed previously in the partially folded TL structure in the presence of a matched ATP analog (*60*). In addition to Asn[479] and Arg[446], Gln[1078] is positioned to engage the 2´-OH group, which could couple the presence of a correctly matched nucleotide in the A site with the closing of the trigger loop. Such architecture could explain the LOF phenotypes observed when mutating this residue in yeast (*57*) consistent with a key role of this residue in NTP/dNTP discrimination, and may provide an explanation on the GOF phenotype of T834P. Preliminary comparison of the TL on-state with the off-state structure of T834P suggests that the NIR of the TL is partially folded and is situated closer to the A site. Thus, the position of Gln[1078] might be critical since on the one hand, it constitutes part of the

nucleotide selection mechanism (Fouqueau et al., 2013; Yuzenkova and Zenkin, 2010), while on the other hand, displacement of Gln^{1078} could trigger extraction of the neighboring Met^{1079} from its hydrophobic pocket, initiating a cascade of events that would lead into the full on state. Our observation of a buried Met^{1079} raises the possibility that T834P may also have a reduced selectivity against dNTP since movement of Gln^{1078} is unlikely to occur while Met^{1079} is buried. However, this is only speculative since residues Asn^{479} and Arg^{446} also contribute to dNTP discrimination. Future experiments will need to be conducted with different substrates, since there is evidence that experimental design influences the architecture of the active site, which may also explain the differences in observed contacts with NTP (*49*, *60*).

Moreover, coupling of BH kink conformation with our observations of the TL suggest that increased activity may be due to favoring of either the pre- or post-translocated state as hypothesized previously (*52*). The Brownian ratchet model suggests that during the NAC an equilibrium between pre- and post-translocation states exist as rearrangements of TL and BH domains occurs for directional movement of polymerase along the DNA (*53-58*). Thus, it is feasible that the conformational landscape is suited for a post-translocation state, which would provide prolonged contacts with the TL and nucleotide for catalysis (*52*, *59*, *130*). Evidence to support this is T834P's ability as a GOF mutant to weakly suppress some lethal LOF TL variants (*198*). However, this will need to be further explored using molecular dynamic simulations and validation with *in vivo* transcription assays.

8.5 ACKNOWLEDGEMENTS

This work was supported by SLAC and Laboratory Directed Research Development grant 0006-15 (A.C.). Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, and Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. We thank Clyde Smith and Craig Ogata for user support at SLAC and APS beamlines, respectively. This work was supported by the National Institute of General Medical Sciences (NIGMS) of the US National Institutes of Health (NIH) under Award numbers GM112686 (G.C), GM112497 (C.O.B), and GM102520 (A. S. B.). G.C acknowledges support from BioXFEL-STC1231306. C.D.K acknowledges support from NIH R01 GM097260 and Welch Foundation Grant A-1763.

9.0 CONCLUSIONS, SIGNIFICANCE, AND FUTURE DIRECTIONS

The work presented in this dissertation has provided a more detailed understanding of Pol II function and methodologies to expand the field of structural biology. Our structure revealing a complete nucleic acid scaffold to high resolution detailed the first complete transcription bubble and upstream DNA duplex bound to Pol II (Figure 10). In addition to DNA/RNA hybrid interactions in the active site, Pol II engages with the minor grooves of upstream/downstream DNA duplex and defines the architecture of the upstream fork of the TB. These interactions gave insight into how an elongating Pol II may function independently during translocation. Since TLB, Rpb5 and Rpb2 regions could be coupled, through extensive observed interactions, it is possible that local TL conformational changes are transferred allosterically to downstream and upstream duplexes to assist global translocation. Furthermore, revelation of a partially folded off-state conformation of the TL, that is stabilized through interactions with the non-template strand and proximal Pol II domains (or allosterically through TFs that stabilize the NTS) increases our understanding of TL dynamics during the NAC. Although speculative this could suggest that TL "off/on" transitions are allosterically coupled to "latch and release" events of the non-template strand phosphate chain by Pol II residues.

To further examine the role, structurally, of the TL and BH during translocation and catalysis, we characterized the GOF T834P variant of Pol II, which biochemically shows a 3-fold increase in elongation rates (*198*). Structural studies of T834P in the presence of a matched

nucleotide revealed the complete TL in the on-state, a conformation only observed previously in two datasets at low resolution (*49*). Our preliminary analysis suggests that kinking of the BH (observed in both off/on state TL structures) potentially shifts equilibrium into the post-translocation state during the NAC, which would prolong interactions of the NIR and tip residues of the TL with the matched nucleotide in the insertion site, explaining T834P's ability to suppress LOF alleles. A surprising revelation of this structure, however, was the observation of Met^{1079} remaining buried during the off/on state transition (Figure 54). Such a conformation may alter our understanding of off/on state TL transitions from previously observed structures. Since on the one hand, the burying of Met^{1079} is seen as a critical step for stabilizing TL off state conformation, but on the other hand, our structures reveal that removal of Met^{1079} from the pocket is not a necessary step for the TL on-state conformation transition.

In the future, characterization of T834P in the context of a full NAS and alternative scaffolds will be necessary, as this may alter NTP positioning in the active site. More importantly, characterizing TL dynamics in the context of other TF which may allosterically play a role in its activity will also be instrumental in our understanding of the NAC. To better understand these interactions, high resolution information is necessary. In this work, we also provided methodologies which will aid in the crystallization of highly relevant, and challenging protein targets, including multi-protein complexes of the transcriptional machinery. Our TEM guided methods are applicable to not only optimizing crystallographic trials through analysis of crystal lattices, but also in assessing samples prior to experiments at XFEL sources. The latter of which has significantly advanced experiments in our lab, by extending the resolution of Pol II complexes beyond the capabilities at the synchrotron due to the high solvent content and radiation sensitivity of the crystals. As such tracing of novel TF bound to Pol II is now possible.

However, improvements in isomorphism are necessary to maintain complete datasets to high resolutions, as well as utilize native zincs for *de novo* phasing of XFEL data.

Nevertheless, such capabilities will undoubtedly allow us to address remaining questions concerning the Pol II nucleotide addition cycle and translocation. For one, what are the roles of the initiation factors, TFIIF and TFIIB, in stabilizing DNA during transcription bubble loading in early transcribing complexes? Moreover, what conformational changes are necessary for TFIIB promoter escape and transitioning of the upstream DNA duplex from contacts with the TFIIB cyclin domains to engage the Rpb2 wedge domain after TFIIB ejection? During early transcribing events, how do TFIIB and TFIIF contribute to the NAS? Do these factors play a role in overcoming promoter pausing of stalled polymerases, and how might they influence TFIIS activity during arrest/abortive transcription at the promoter? After escape, what changes occur that contribute to NTS stability when TFIIF is replaced by the elongation factor Spt4/5, and again, how might Spt4/5 allosterically affect the NAC and TL dynamics?

We could gain further insight into TL dynamics by using time-resolved SFX experiments of the T834P variant through the addition of a matched nucleotide. The recent advancements in mixing jet technologies will allow for the observation of off/on state transitions of the TL. If successful, this would greatly enhance our understanding of the NAC, as we could potentially examine pre- and post-translocation states of an elongating Pol II.

10.0 RELEVANT PUBLICATIONS

1. **Barnes CO*,** Kovaleva EG*, Fu X, Stevenson HP, Brewster AS, DePonte DP, Baxter EL, Cohen AE and Calero G. (2016) Assessment of microcrystal quality by transmission electron microscopy for efficient serial femtosecond crystallography. *Arch Biochem Biophys.* **602, 61-68.**

2. Stevenson HP*, Lin G*, **Barnes CO***, Sutkeviciute I, Krzysiak T, Weiss SC, et al. (2016) Applications of TEM for the evaluation and optimization of crystal growth. *Acta Cryst.* **D72(5), 603-15.** ***(Cover Article)***

3. Baxter EL, Aguila L, Alonso-Mori R, **Barnes CO**, Bonagura CA, Brehmer W, Brunger AT, *et al*. (2016) High-density grids for efficient data collection from multiple crystals. *Acta Cryst*, **D72**, 2-11.

4. **Barnes CO,** Calero M, Malik I, Graham B , Saphr H, Cohen A, Guowu L, Brown I, Zhang Q, Pullara F, Trakeselis M, Kaplan C, and Calero G. (2015) Crystal structure of a transcribing RNA Polymerase II complex reveals a complete transcription bubble. *Molecular Cell*, ***59****(2)*, 258-269.

5. Gonzalez A, Aguila L, Alonso-Mori R, **Barnes CO**, Baxter E, Brehmer W, et al. (2014) Goniometer-based femtosecond crystallography with X-ray free electron lasers. *PNAS*, **111**(48), 17122-7.

6. Stevenson HP, Makhov AM, Calero M, Mathews I, Lin G, **Barnes CO**, Santamaria H, Ross T, Soitis M, Khosla C, Nagaragan V, Conway JF, Cohen A, and Calero G. (2014) Use of transmission electron microscopy to identify nanocrystals of challenging protein targets. *PNAS*, **111**(23): 8470-5.

In preparation

7. **Barnes CO**, Wu Y, Song J, Lin G, Baxter EL, Brewster AS, Nagarajan V, Holmes A, Soltis M, Ahn J, Cohen AE, Calero G. (2016) Structural basis of enzymatic specificity and inhibition of the dGTP triphosphohydrolase from *Escherichia coli* revealed by XFEL and chemical cross-linking.

APPENDIX A

BIOCHEMICAL VALIDATION OF RPB2 WEDGE

Biochemical validation and analysis of Rpb2 wedge interactions (see **Chapter 3**) were performed in collaboration with Dr. Indranil Malik from the lab of Dr. Craig Kaplan. Dr. Malik was responsible for the following experimental methods:

In vitro elongation assay

Pol II enzymes for in vitro assays were purified from yeast strains expressing wild type or mutant *rpb2* gene from a low copy plasmid, as previously described above. *In vitro* elongation assays were performed as described in (*56, 57*) with minor modifications in the amount of nucleic acids and Pol II used for elongation complexes. In brief, an RNA-DNA hybrid was formed by annealing 20 pmol of the 81 nt template oligo, CKO223, with an equimolar amount of a γ-^{32}P-ATP end-labeled 9-mer RNA (RNA9), in 10 µl of *in vitro* transcription buffer (TB, 20 mM Tris pH 8.0, 40 mM KCl, 10 mM MgCl2, 10% glycerol, 0.25 mg/ml bovine serum albumin, 2 mM DTT). 2.5 µl hybrid were incubated with 2.5 µg of Pol II (2.5 µl) at room temperature for 5 minutes followed by an incubation with 40 pmol of non-template oligo, CKO222, for 5 minutes at room temperature. Resulting complexes were diluted in 1X TB and used for time course

reactions as described (*57*). Time course products were resolved in 13.5% acrylamide-bisacrylamide (19:1) gel containing 1X TBE and 7M urea. Run-off transcripts were quantified as fraction of total signal per lane for transcripts 10 nucleotides and above, using Image Lab (Bio-Rad) software. Half-times and maximum elongation rates were determined using GraphPad Prism 5.0, as described previously in (*56*).

Chromatin Immunoprecipitation (ChIP) assays

Chromatin immunoprecipitation experiments for *in vivo* elongation rate determination were performed as described previously in (*253*), with slight modifications. Epitope-tagged (*RPB3::3XFLAG::kanMX*) wild type or mutant strain containing a galactose inducible *YLR454w* reporter (*kanMX::GAL1p::YLR454w*) gene were used for ChIP assays. Overnight grown cultures were used to inoculate larger cultures of YP-Galactose (1% yeast extract, 2% bacto-peptone, 2% galactose) to amplify. At mid-log phase, pre- glucose shut-off samples were taken for the 0 minute time point and cross-linked with formaldehyde (1% final, v/v), followed by the addition of glucose (4% final) to shut-off transcription. Aliquots were removed after 2, 4, 6 and 8 minutes time points and cross-linked as above. Chromatin preparations, followed by immunoprecipitations with anti-FLAG M2 beads (M8823, SIGMA) were done as in (*57*), with the exception of using Bioruptor 300, Diagenode (manufacturer guidelines) for shearing cross-linked chromatin. Input or immunoprecipitated DNAs were analyzed by quantitative PCR using CFX96 (Bio-Rad). A non-transcribed telomeric region was used for normalization to obtain fold-enrichment values (*253*).

Northern blotting and primer extension analysis

Northern blotting was performed as previously described (*56*) essentially following the instructions of GeneScreen hybridization membranes (Perkin-Elmer) with minor modifications. Twenty micrograms total yeast RNA purified as described (*254*) was prepared in Glyoxal sample load dye (Ambion). Prehybridization solution was made without SSC or SSPE buffers, but contained 50% formamide, 5X Denhardt's solution, 10% Dextran sulfate, 1M NaCl, 50mM Tris-HCl pH7.5, 0.1% sodium pyrophosphate, 0.1% SDS and 500 µg/ml denatured salmon sperm DNA. Probes for northern blots were radiolabeled with ^{32}P-dATP using the Decaprime II kit (Ambion) according to manufacturer's instructions. Blots were washed twice each- in 2x SSC for 15 minutes at 42˚C, in 5X SSC with 0.5% SDS for 30 minutes at 65˚C, and in 0.2x SSC for 30 minutes at room temperature. Primer extension assays were performed as described (*56*), which is a modified version of (*255*). 30 µg total yeast RNA was annealed with ^{32}P end-labeled oligo in a 15 µl reaction volume. Reverse transcription reaction was done by M-MLV Reverse Transcriptase (Fermentas) supplemented with RNase Inhibitor (Fermentas), in a total reaction volume of 45 µl. Extension products were separated on 8% polyacrylamide gels (19:1 acrylamide:bisacrylamide, Bio-Rad) containing 1X TBE and 7M urea. Northern blots and primer extension gels were visualized by phosphorimaging (Bio-Rad or GE Healthcare) and quantified using Quantity One (Bio-Rad) and ImageQuant 5.0 (GE), respectively. Primers and probe sequences are available upon request.

APPENDIX B

HOMOPROTOCATECHUATE 2,3-DIOXYGENASE EXPERIMENTS

Data presented in Chapter 7 was done in tandem with Dr. Elena Kovaleva in collaboration with the lab of Dr. Aina Cohen. Dr. Kovaleva was responsible for 2,3 – HPCD sample preparation, data collection, and processing as described below:

Enzyme purification and crystallization

Recombinant homoprotocatechuate 2,3-dioxygenase from *B. fuscum* was expressed in *E.coli* and purified using procedures described previously (*256*, *257*). The enzyme preparations used in this study included a wild type (WT) enzyme and an active site variant with His^{200} substituted for Gln (H200Q). Previously published crystallization conditions and procedures that produce diffraction quality macro-crystals of 2,3-HPCD enzymes (*258*) were modified for injector-based experiments to scale up production of micro-crystal slurries using an over-nucleation approach. Protein solutions (7 – 10 mg/ml) were gently mixed in a 1:1 ratio with crystallization solution consisting of 12 – 14% PEG6000, 0.1 M calcium chloride, and 0.1 M MES pH 5.8, ensuring that no precipitate forms at these concentrations of components. After a few minutes of equilibration, a 20 µl aliquot of the 1:1 stock mixture (protein and crystallization solution) was seeded with a

few roughly crushed macro-crystals in an eppendorf tube, and growth of many needle-like crystals was observed after a few hours of incubation at 20 °C. Every 2 – 4 hours 50-200 µL of a fresh mixture of protein and crystallization solution was added into the crystallization tubes to maintain maximal growth of needle-like crystals, followed by a final 12 hour period of growth to achieve saturation in a total of 2 – 3 ml volume. The WT enzyme preparation produced slightly larger population of initial crystals (thin rods), whereas the H200Q variant preparation produced slurries with needle-like crystals.

2,3 – HPCD sample preparation for liquid-jet SFX experiments

A mixture of glass beads (0.1 and 0.5 mm) was used to crush the crystals to achieve a smaller and more uniform sample size for injector-based studies. The WT sample, consisting of thin rods, was vortexed with the glass beads mixture several times at medium to high speed for 60 seconds, followed by a 10-minute rest period. Slurries of needle-like crystals of the H200Q variant were concentrated by centrifugation at 2,000 rpm for 5 to 10 minutes prior to the addition of the glass beads mixture and additional centrifugation for 2 minutes. The efficiency of crystal crushing using either the vortexing or microcentrifugation methods were monitored qualitatively by light microscopy, and by visually observing the light scattering of sample medium using a flashlight. Crushed samples were passed through a 20 µm filter to remove particulates and larger crystalline material prior to loading into the sample reservoir for diffraction studies.

2,3 – HPCD liquid-jet SFX data collection, processing and refinement

Diffraction data were collected at the CXI end station of LCLS using 8.8 keV X-ray pulses with 40 fs duration and 1 μm beam focus at the X-ray interaction point. Sample delivery using gas dynamic virtual nozzle (GDVN) was carried out at a flow rate of 20 μl/sec at 16 °C. Diffraction images were collected at 120 Hz using a CSPAD detector and processed using *ccbtx.xfel* software package (*195*, *226*). The WT enzyme sample resulted in 814 indexable images during 6.7 min of data collection. A total of 15.6 min of data collection using the H200Q variant sample resulted in a dataset based on 14026 merged images. During indexing, secondary lattices indicative of multiple crystals were identified by first indexing, then removing primary lattices and then attempting to index the remaining bright reflections (*195*, *226*).

BIBLIOGRAPHY

1. J. D. Watson, F. H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. **171**, 737–738 (1953).

2. J. D. Watson, F. H. Crick, Genetical implications of the structure of deoxyribonucleic acid. *Nature*. **171**, 964–967 (1953).

3. F. Crick, Central dogma of molecular biology. *Nature* (1970), doi:10.1038/227561a0.

4. J. M. Coffin, H. Fan, The Discovery of Reverse Transcriptase. *Annu Rev Virol*. **3**, 29–51 (2016).

5. S. Geisler, J. Coller, RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature Reviews Molecular Cell Biology*. **14**, 699–712 (2013).

6. V. Ambros, The functions of animal microRNAs. *Nature*. **431**, 350–355 (2004).

7. D. P. Bartel, MicroRNAs. *Cell*. **116**, 281–297 (2004).

8. A. Vannini, P. Cramer, Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Molecular Cell*. **45**, 439–446 (2012).

9. P. A. Konstantinopoulos, A. G. Papavassiliou, Seeing the future of cancer-associated transcription factor drug targets. *JAMA*. **305**, 2349–2350 (2011).

10. C. Villicaña, G. Cruz, M. Zurita, The basal transcription machinery as a target for cancer therapy. *Cancer Cell Int.* **14**, 18 (2014).

11. X. Tong, F. Wang, C. J. Thut, E. Kieff, The Epstein-Barr virus nuclear protein 2 acidic domain can interact with TFIIB, TAF40, and RPA70 but not with TATA-binding protein. *J. Virol.* **69**, 585–588 (1995).

12. I. Agostini *et al.*, The HIV-1 Vpr co-activator induces a conformational change in TFIIB. *FEBS Lett.* **450**, 235–239 (1999).

13. Y. Yang *et al.*, HBV X protein (HBX) interacts with general transcription factor TFIIB both in vitro and in vivo. *Chin. Med. Sci. J.* **14**, 152–157 (1999).

14. L. Li *et al.*, General transcription factor IIb overexpression and a potential link to proliferation in human hepatocellular carcinoma. *Pathol. Oncol. Res.* **19**, 195–203 (2013).

15. M. Lu *et al.*, Overexpression of TFIIB-related factor 2 is significantly correlated with tumor angiogenesis and poor survival in patients with esophageal squamous cell cancer. *Med. Oncol.* **30**, 553 (2013).

16. I. Veras, E. M. Rosen, L. Schramm, Inhibition of RNA polymerase III transcription by BRCA1. *Journal of Molecular Biology*. **387**, 523–531 (2009).

17. M. Lu *et al.*, TFIIB-related factor 2 over expression is a prognosis marker for early-stage non-small cell lung cancer correlated with tumor angiogenesis. *PLoS ONE*. **9**, e88032 (2014).

18. Y. Tian *et al.*, TFIIB-related factor 2 is associated with poor prognosis of nonsmall cell lung cancer patients through promoting tumor epithelial-mesenchymal transition. *Biomed Res Int*. **2014**, 530786–13 (2014).

19. B. L. Allen, D. J. Taatjes, The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology*. **16**, 155–166 (2015).

20. L. T. Vassilev *et al.*, In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*. **303**, 844–848 (2004).

21. H. M. Berman, G. J. Kleywegt, H. Nakamura, J. L. Markley, The future of the protein data bank. *Biopolymers*. **99**, 218–222 (2012).

22. R. Sousa, Y. J. Chung, J. P. Rose, B. C. Wang, Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature*. **364**, 593–599 (1993).

23. G. M. Cheetham, T. A. Steitz, Structure of a transcribing T7 RNA polymerase initiation complex. *Science*. **286**, 2305–2309 (1999).

24. G. Zhang *et al.*, Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution. *Cell*. **98**, 811–824 (1999).

25. P. Cramer, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*. **292**, 1863–1876 (2001).

26. R. G. Roeder, W. J. Rutter, Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*. **224**, 234–237 (1969).

27. L. B. Schwartz, V. E. Sklar, J. A. Jaehning, R. Weinmann, R. G. Roeder, Isolation and partial characterization of the multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in the mouse myeloma, MOPC 315. *Journal of Biological Chemistry*. **249**, 5889–5897 (1974).

28. N. F. Lue, R. D. Kornberg, Accurate initiation at RNA polymerase II promoters in extracts from Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*. **84**, 8839–8843 (1987).

29. A. M. Edwards *et al.*, Purification and lipid-layer crystallization of yeast RNA polymerase II. *Proceedings of the National Academy of Sciences*. **87**, 2122–2126 (1990).

30. A. M. Edwards, C. M. Kane, R. A. Young, R. D. Kornberg, Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J. Biol. Chem.* **266** (1991).

31. M. Choder, R. A. Young, A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Molecular and Cellular Biology*. **13**, 6984–6991 (1993).

32. K. McKune, K. L. Richards, A. M. Edwards, R. A. Young, N. A. Woychik, RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. *Yeast*. **9**, 295–299 (1993).

33. F. J. Asturias, G. D. Meredith, C. L. Poglitsch, R. D. Kornberg, Two conformations of RNA polymerase II revealed by electron crystallography. *Journal of Molecular Biology*. **272**, 536–540 (1997).

34. F. J. Asturias, R. D. Kornberg, A Novel Method for Transfer of Two-Dimensional Crystals from the Air/Water Interfacé to Specimen Grids. *Journal of Structural Biology*. **114**, 60–66 (1995).

35. L. C. Myers, K. Leuther, D. A. Bushnell, C. M. Gustafsson, R. D. Kornberg, Yeast RNA polymerase II transcription reconstituted with purified proteins. *Methods*. **12**, 212–216 (1997).

36. J. Fu *et al.*, Yeast RNA Polymerase II at 5 Å Resolution. *Cell*. **98**, 799–810 (1999).

37. P. Cramer *et al.*, Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*. **288**, 640–649 (2000).

38. A. L. Gnatt, Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 A Resolution. *Science*. **292**, 1876–1882 (2001).

39. P. Cramer, Deciphering the RNA polymerase II structure: a personal perspective. *Nat. Struct. Mol. Biol.* **13**, 1042–1044 (2006).

40. R. D. Kornberg, The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences*. **104**, 12955–12961 (2007).

41. D. A. Bushnell, R. D. Kornberg, Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proceedings of the National Academy of Sciences*. **100**, 6969–6973 (2003).

42. K.-J. Armache, H. Kettenberger, P. Cramer, Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences*. **100**, 6964–6968 (2003).

43. K. D. Westover, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*. **303**, 1014–1016 (2004).

44. K. D. Westover, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*. **119**, 481–489 (2004).

45. H. Kettenberger, K.-J. Armache, P. Cramer, Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. *Molecular Cell*. **16**, 955–965 (2004).

46. T. A. Steitz, A mechanism for all polymerases. *Nature*. **391**, 231–232 (1998).

47. D. G. Vassylyev, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, I. Artsimovitch, Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*. **448**, 157–162 (2007).

48. A. C. M. Cheung, P. Cramer, Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*. **471**, 249–253 (2011).

49. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, R. D. Kornberg, Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell*. **127**, 941–954 (2006).

50. D. Temiakov *et al.*, Structural Basis of Transcription Inhibition by Antibiotic Streptolydigin. *Molecular Cell*. **19**, 655–666 (2005).

51. S. Tuske *et al.*, Inhibition of Bacterial RNA Polymerase by Streptolydigin: Stabilization of a Straight-Bridge-Helix Active-Center Conformation. *Cell*. **122**, 541–552 (2005).

52. L. Tan, S. Wiesler, D. Trzaska, H. C. Carney, R. O. Weinzierl, Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J Biol*. **7**, 40 (2008).

53. G. Bar-Nahum *et al.*, A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell*. **120**, 183–193 (2005).

54. V. Epshtein *et al.*, Swing-Gate Model of Nucleotide Entry into the RNA Polymerase Active Center. *Molecular Cell*. **10**, 623–634 (2002).

55. T. Fouqueau, M. E. Zeller, A. C. Cheung, P. Cramer, M. Thomm, The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Research*. **41**, 7048–7059 (2013).

56. C. D. Kaplan, H. Jin, I. L. Zhang, A. Belyanin, Dissection of Pol II trigger loop function

and Pol II activity-dependent control of start site selection in vivo. *PLoS Genet.* **8**, e1002627 (2012).

57. C. D. Kaplan, K.-M. Larsson, R. D. Kornberg, The RNA Polymerase II Trigger Loop Functions in Substrate Selection and Is Directly Targeted by α-Amanitin. *Molecular Cell*. **30**, 547–556 (2008).

58. M. L. Kireeva *et al.*, Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase. *BMC biophysics*. **5**, 11 (2012).

59. F. Brueckner, P. Cramer, Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation. *Nat. Struct. Mol. Biol.* **15**, 811–818 (2008).

60. A. C. M. Cheung, S. Sainsbury, P. Cramer, Structural basis of initial RNA polymerase II transcription. *The EMBO Journal*. **30**, 4755–4763 (2011).

61. H. Kettenberger, K.-J. Armache, P. Cramer, Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*. **114**, 347–357 (2003).

62. D. Wang *et al.*, Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science*. **324**, 1203–1206 (2009).

63. J. F. Sydow *et al.*, Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Molecular Cell*. **34**, 710–721 (2009).

64. G. E. Damsma, A. Alt, F. Brueckner, T. Carell, P. Cramer, Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nat. Struct. Mol. Biol.* **14**, 1127–1133 (2007).

65. F. Brueckner, U. Hennecke, T. Carell, P. Cramer, CPD damage recognition by transcribing RNA polymerase II. *Science*. **315**, 859–862 (2007).

66. C. Walmacq *et al.*, Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage. *Molecular Cell*. **46**, 18–29 (2012).

67. R. Landick, The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* **34**, 1062–1066 (2006).

68. R. Landick, Transcriptional pausing without backtracking. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8797–8798 (2009).

69. F. Toulmé *et al.*, GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *The EMBO Journal*. **19**, 6853–6859 (2000).

70. C. M. Tyree *et al.*, Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes Dev.* **7**, 1254–1265 (1993).

71. W. Deng, S. G. E. Roberts, TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma*. **116**, 417–429 (2007).

72. J. D. Parvin, P. A. Sharp, DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell*. **73**, 533–540 (1993).

73. D. B. Nikolov *et al.*, Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*. **377**, 119–128 (1995).

74. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science*. **303**, 983–988 (2004).

75. O. Littlefield, Y. Korkhin, P. B. Sigler, The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proceedings of the National Academy of Sciences*. **96**, 13668–13673 (1999).

76. S. D. Bell, C. P. Magill, S. P. Jackson, Basal and regulated transcription in Archaea. *Biochem. Soc. Trans.* **29**, 392–395 (2001).

77. D. G. Vassylyev *et al.*, Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*. **417**, 712–719 (2002).

78. H.-T. Chen, S. Hahn, Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell*. **119**, 169–180 (2004).

79. D. Kostrewa *et al.*, RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*. **462**, 323–330 (2009).

80. X. Liu, D. A. Bushnell, D. Wang, G. Calero, R. D. Kornberg, Structure of an RNA Polymerase II-TFIIB Complex and the Transcription Initiation Mechanism. *Science*. **327**, 206–209 (2010).

81. S. Sainsbury, J. Niesser, P. Cramer, Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*. **493**, 1–5 (2012).

82. M. Pal, A. S. Ponticelli, D. S. Luse, The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Molecular Cell*. **19**, 101–110 (2005).

83. P. G. Rani, J. A. Ranish, S. Hahn, RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Molecular and Cellular Biology*. **24**, 1709–1720 (2004).

84. C. Plaschka *et al.*, Transcription initiation complex structures elucidate DNA opening. *Nature*. **533**, 353–358 (2016).

85. Y. He, J. Fang, D. J. Taatjes, E. Nogales, Structural visualization of key steps in human

transcription initiation. *Nature*. **495**, 481–486 (2013).

86. K. Murakami *et al.*, Formation and Fate of a Complete 31-Protein RNA Polymerase II Transcription Preinitiation Complex. *Journal of Biological Chemistry*. **288**, 6325–6332 (2013).

87. Z. A. Chen *et al.*, Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal*. **29**, 717–726 (2010).

88. K. Murakami *et al.*, Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex. *Science*. **342**, 1238724–1238724 (2013).

89. C. Plaschka *et al.*, Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*. **518**, 376–380 (2015).

90. K. Murakami *et al.*, Structure of an RNA polymerase II preinitiation complex. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13543–13548 (2015).

91. R. K. Louder *et al.*, Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature*. **531**, 604–609 (2016).

92. Y. He *et al.*, Near-atomic resolution visualization of human transcription promoter opening. *Nature*. **533**, 359–365 (2016).

93. P. J. Robinson *et al.*, Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell*. **166**, 1411–1422.e16 (2016).

94. P. Cramer, Structure determination of transient transcription complexes. *Biochem. Soc. Trans.* **44**, 1177–1182 (2016).

95. E. Nogales, R. K. Louder, Y. He, Cryo-EM in the study of challenging systems: the human transcription pre-initiation complex. *Current Opinion in Structural Biology*. **40**, 120–127 (2016).

96. U. Fiedler, H. T. Timmers, Analysis of the open region of RNA polymerase II transcription complexes in the early phase of elongation. *Nucleic Acids Research*. **29**, 2706–2714 (2001).

97. N. Naryshkin, A. Revyakin, Y. Kim, V. Mekler, R. H. Ebright, Structural organization of the RNA polymerase-promoter open complex. *Cell*. **101**, 601–611 (2000).

98. A. N. Kapanidis *et al.*, Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science*. **314**, 1144–1147 (2006).

99. A. Revyakin, C. Liu, R. H. Ebright, T. R. Strick, Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science*. **314**, 1139–1143 (2006).

100. W.-H. Chung *et al.*, RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Molecular Cell*. **12**, 1003–1013 (2003).

101. F. Pullara *et al.*, A general path for large-scale solubilization of cellular proteins: from membrane receptors to multiprotein complexes. *Protein Expression and Purification*. **87**, 111–119 (2013).

102. T. T. Richardson *et al.*, Unwinding of primer-templates by archaeal family-B DNA polymerases in response to template-strand uracil. *Nucleic Acids Research*. **41**, 2466–2478 (2013).

103. P. A. Meyer, P. Ye, M.-H. Suh, M. Zhang, J. Fu, Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data. *Journal of Biological Chemistry*. **284**, 12933–12939 (2009).

104. A. Vagin, A. Teplyakov, An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1622–1624 (2000).

105. E. Blanc *et al.*, Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221 (2004).

106. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

107. M. D. Winn, G. N. Murshudov, M. Z. Papiz, Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol*. **374**, 300–321 (2003).

108. A. T. Brunger *et al.*, Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).

109. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

110. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

111. B. DeLaBarre, A. T. Brunger, Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 923–932 (2006).

112. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*. **98**, 10037–10041 (2001).

113. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science*. **336**, 1030–1033 (2012).

114. K. Schwinghammer *et al.*, Structure of human mitochondrial RNA polymerase elongation complex. *Nat. Struct. Mol. Biol.* **20**, 1–7 (2013).

115. A. Hirata, B. J. Klein, K. S. Murakami, The X-ray crystal structure of RNA polymerase from Archaea. *Nature*. **451**, 851–854 (2008).

116. H. Spahr, G. Calero, D. A. Bushnell, R. D. Kornberg, Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9185–9190 (2009).

117. C. Engel, S. Sainsbury, A. C. Cheung, D. Kostrewa, P. Cramer, RNA polymerase I structure and transcription regulation. *Nature*. **502**, 1–19 (2013).

118. C. D. Kaplan, Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae. *Biochim Biophys Acta*. **1829**, 39–54 (2013).

119. B. Treutlein *et al.*, Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Molecular Cell*. **46**, 136–146 (2012).

120. E. Kashkina *et al.*, Multisubunit RNA Polymerases Melt Only a Single DNA Base Pair Downstream of the Active Site. *Journal of Biological Chemistry*. **282**, 21578–21582 (2007).

121. N. Korzheva *et al.*, A structural model of transcription elongation. *Science*. **289**, 619–625 (2000).

122. C. Liu, C. T. Martin, Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. *Journal of Molecular Biology*. **308**, 465–475 (2001).

123. J. T. Stivers, 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Research*. **26**, 3837–3844 (1998).

124. X. Liu, D. A. Bushnell, D. A. Silva, X. Huang, R. D. Kornberg, Initiation Complex Structure and Promoter Proofreading. *Science*. **333**, 633–637 (2011).

125. M. Feig, Z. F. Burton, RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation. *Biophys. J.* **99**, 2577–2586 (2010).

126. M. H. Larson *et al.*, Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6555–6560 (2012).

127. H. Braberg *et al.*, From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell*. **154**, 775–788 (2013).

128. F. Malagon *et al.*, Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics*. **172**, 2201–2209 (2006).

129. Y. Harada *et al.*, Direct observation of DNA rotation during transcription by Escherichia coli RNA polymerase. *Nature*. **409**, 113–115 (2001).

130. F. W. Martinez-Rucobo, S. Sainsbury, A. C. Cheung, P. Cramer, Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. *The EMBO Journal*. **30**, 1302–1310 (2011).

131. B. J. Klein *et al.*, RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 546–550 (2011).

132. A. Sevostyanova, I. Artsimovitch, Functional analysis of Thermus thermophilus transcription factor NusG. *Nucleic Acids Research*. **38**, 7432–7445 (2010).

133. Y. Zhang *et al.*, Structural Basis of Transcription Initiation. *Science*. **338**, 1076–1080 (2012).

134. Y. Zuo, T. A. Steitz, Crystal structures of the E. coli transcription initiation complexes with a complete bubble. *Molecular Cell*. **58**, 534–540 (2015).

135. Y. Yuzenkova, N. Zenkin, Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10878–10883 (2010).

136. S. Dutta, H. M. Berman, Large macromolecular complexes in the Protein Data Bank: a status report. *Structure/Folding and Design*. **13**, 381–388 (2005).

137. E. M. Marcotte *et al.*, Detecting protein function and protein-protein interactions from genome sequences. *Science*. **285**, 751–753 (1999).

138. Y.-C. Han *et al.*, Prediction and characterization of protein-protein interaction network in Bacillus licheniformis WX-02. *Sci Rep*. **6**, 19486 (2016).

139. E. D. Levy, J. B. Pereira-Leal, C. Chothia, S. A. Teichmann, 3D complex: a structural classification of protein complexes. *PLoS Comput Biol*. **2**, e155 (2006).

140. P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*. **293**, 321–331 (1999).

141. S. Vucetic, C. J. Brown, A. K. Dunker, Z. Obradovic, Flavors of protein disorder. *Proteins*. **52**, 573–584 (2003).

142. W. N. Price *et al.*, Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol*. **27**, 51–57 (2009).

143. S. Radaev, S. Li, P. D. Sun, A survey of protein-protein complex crystallizations. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 605–612 (2006).

144. E. P. Carpenter, K. Beis, A. D. Cameron, S. Iwata, Overcoming the challenges of

membrane protein crystallography. *Current Opinion in Structural Biology*. **18**, 581–586 (2008).

145. E. I. Mokhonova, V. V. Mokhonov, H. Akama, T. Nakae, Forceful large-scale expression of “problematic” membrane proteins. *Biochem. Biophys. Res. Commun.* **327**, 650–655 (2005).

146. F. Saida, M. Uzan, B. Odaert, F. Bontems, Expression of highly toxic genes in E. coli: special strategies and genetic tools. *Curr. Protein Pept. Sci.* **7**, 47–56 (2006).

147. A. Ashkenazi *et al.*, Mapping the Cd4 Binding-Site for Human-Immunodeficiency-Virus by Alanine-Scanning Mutagenesis. *Proceedings of the National Academy of Sciences*. **87**, 7150–7154 (1990).

148. M. Blaber, W. A. Baase, N. Gassner, B. W. Matthews, Alanine scanning mutagenesis of the alpha-helix 115-123 of phage T4 lysozyme: effects on structure, stability and the binding of solvent. *Journal of Molecular Biology*. **246**, 317–330 (1995).

149. B. C. Cunningham, J. A. Wells, High-Resolution Epitope Mapping of Hgh-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science*. **244**, 1081–1085 (1989).

150. C. S. Gibbs, M. J. Zoller, Identification of Electrostatic Interactions That Determine the Phosphorylation Site Specificity of the Camp-Dependent Protein-Kinase. *Biochemistry-Us*. **30**, 5329–5334 (1991).

151. V. Cherezov *et al.*, High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*. **318**, 1258–1265 (2007).

152. F. Dupeux, M. Rower, G. Seroul, D. Blot, J. A. Marquez, A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 915–919 (2011).

153. T. Hino *et al.*, G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature*. **482**, 237–240 (2012).

154. S. G. F. Rasmussen *et al.*, Structure of a nanobody-stabilized active state of the β(2) adrenoceptor. *Nature*. **469**, 175–180 (2011).

155. J. Steyaert, B. K. Kobilka, Nanobody stabilization of G protein-coupled receptor conformational states. *Curr. Opin. Struct. Biol.* **21**, 567–572 (2011).

156. T. Bergfors, Seeds to crystals. *Journal of Structural Biology*. **142**, 66–76 (2003).

157. A. McPherson, The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. *Methods of biochemical analysis*. **23**, 249–345 (1976).

158. E. A. Stura, I. A. Wilson, Applications of the Streak Seeding Technique in Protein

Crystallization. *J Cryst Growth*. **110**, 270–282 (1991).

159. A. D'Arcy, T. Bergfors, S. W. Cowan-Jacob, M. Marsh, Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta crystallographica. Section F, Structural biology communications*. **70**, 1117–1126 (2014).

160. G. C. Ireton, B. L. Stoddard, Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 601–605 (2004).

161. T. Stuwe *et al.*, Architecture of the fungal nuclear pore inner ring complex. *Science*. **350**, 56–64 (2015).

162. N. Zhong *et al.*, Optimizing Production of Antigens and Fabs in the Context of Generating Recombinant Antibodies to Human Proteins. *PLoS ONE*. **10**, e0139695 (2015).

163. M. Paduch *et al.*, Generating conformation-specific synthetic antibodies to trap proteins in selected functional states. *Methods*. **60**, 3–14 (2013).

164. A. K. Shukla *et al.*, Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature*. **497**, 137–141 (2013).

165. D. J. Kissick, D. Wanapun, G. J. Simpson, Second-order nonlinear optical imaging of chiral crystals. *Annual review of analytical chemistry (Palo Alto, Calif.)*. **4**, 419–437 (2011).

166. S. Desbois, S. A. Seabrook, J. Newman, Some practical guidelines for UV imaging in the protein crystallization laboratory. *Acta crystallographica. Section F, Structural biology and crystallization communications*. **69**, 201–208 (2013).

167. G. Calero, A. E. Cohen, J. R. Luft, J. Newman, E. H. Snell, Identifying, studying and making good use of macromolecular crystals. *Acta crystallographica. Section F, Structural biology communications*. **70**, 993–1008 (2014).

168. H. P. Stevenson *et al.*, Use of transmission electron microscopy to identify nanocrystals of challenging protein targets. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8470–8475 (2014).

169. H. N. Chapman *et al.*, Femtosecond X-ray protein nanocrystallography. *Nature*. **470**, 73–77 (2011).

170. Y. Kang *et al.*, Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature*. **523**, 561–567 (2015).

171. J. A. Rodriguez *et al.*, Structure of the toxic core of α-synuclein from invisible crystals. *Nature*. **525**, 486–490 (2015).

172. D. Shi, B. L. Nannenga, M. G. Iadanza, T. Gonen, Three-dimensional electron crystallography of protein microcrystals. *eLife*. **2**, 213–17 (2013).

173. J. Tenboer *et al.*, Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science*. **346**, 1242–1246 (2014).

174. G. Calero *et al.*, Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. *Nat Struct Biol*. **9**, 912–917 (2002).

175. H. P. Stevenson *et al.*, Transmission electron microscopy as a tool for nanocrystal characterization pre- and post-injector. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. **369**, 20130322–20130322 (2014).

176. K. De Vos, Cell Counter. *rsb.info.nih.gov* (2001), (available at http://rsb.info.nih.gov/ij/plugins/cell-counter.html).

177. A. J. McCoy *et al.*, Phaser crystallographic software. *Journal of applied crystallography*. **40**, 658–674 (2007).

178. M. C. Hilgart *et al.*, Automated sample-scanning methods for radiation damage mitigation and diffraction-based centering of macromolecular crystals. *Journal of synchrotron radiation*. **18**, 717–722 (2011).

179. L. Redecke *et al.*, Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science*. **339**, 227–230 (2013).

180. B. G. Abdallah, S. Roy-Chowdhury, J. Coe, P. Fromme, A. Ros, High throughput protein nanocrystal fractionation in a microfluidic sorter. *Anal Chem*. **87**, 4159–4167 (2015).

181. C. Kupitz *et al.*, Microcrystallization techniques for serial femtosecond crystallography using photosystem II from Thermosynechococcus elongatus as a model system. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. **369**, 20130316–20130316 (2014).

182. A. McPherson, A. J. Malkin, G. Kuznetsov Yu, Atomic force microscopy in the study of macromolecular crystal growth. *Annual review of biophysics and biomolecular structure*. **29**, 361–410 (2000).

183. M. Sleutel, A. E. Van Driessche, Role of clusters in nonclassical nucleation and growth of protein crystals. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E546–53 (2014).

184. B. Heras, J. L. Martin, Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 1173–1180 (2005).

185. R. S. Brown, C. Ferguson, A. Kingswell, F. K. Winkler, K. R. Leonard, Electron-Microscopic Study of Crystals of the Xenopus-Laevis Transcription Factor-Iiia-5s Ribosomal-Rna Complex. *Proceedings of the National Academy of Sciences*. **85**, 3802–3804 (1988).

186. C. E. Conrad *et al.*, A novel inert crystal delivery medium for serial femtosecond

crystallography. *IUCrJ*. **2**, 421–430 (2015).

187. R. Fromme *et al.*, Serial femtosecond crystallography of soluble proteins in lipidic cubic phase. *IUCrJ*. **2**, 545–551 (2015).

188. M. Ibrahim *et al.*, Improvements in serial femtosecond crystallography of photosystem II by optimizing crystal uniformity using microseeding procedures. *Structural Dynamics*. **2**, 041705–041708 (2015).

189. Y. Ding *et al.*, Measurements and simulations of ultralow emittance and ultrashort electron beams in the linac coherent light source. *Phys. Rev. Lett.* **102**, 254801 (2009).

190. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature*. **406**, 752–757 (2000).

191. S. Boutet *et al.*, High-resolution protein structure determination by serial femtosecond crystallography. *Science*. **337**, 362–364 (2012).

192. R. A. Kirian *et al.*, Femtosecond protein nanocrystallography-data analysis methods. *Opt Express*. **18**, 5713–5723 (2010).

193. T. A. White *et al.*, Crystallographic data processing for free-electron laser sources. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1231–1240 (2013).

194. H. M. Ginn *et al.*, Structure of CPV17 polyhedrin determined by the improved analysis of serial femtosecond crystallographic data. *Nat Comms*. **6**, 6435 (2015).

195. J. Hattne *et al.*, Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nat. Methods*. **11**, 545–548 (2014).

196. W. Kabsch, Processing of X-ray snapshots from crystals in random orientations. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2204–2216 (2014).

197. A. Y. Lyubimov *et al.*, Advances in X-ray free electron laser (XFEL) diffraction data processing applied to the crystal structure of the synaptotagmin-1 / SNARE complex. *eLife*. **5**, 213 (2016).

198. C. Qiu, O. C. Erinne, J. Dave, P. Cui, H. Jin, High-resolution phenotypic landscape of the RNA Polymerase II trigger loop. *bioRxiv* (2016), doi:10.1101/068726.

199. G. D. Calvey, A. M. Katz, C. B. Schaffer, L. Pollack, Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers. *Struct Dyn*. **3**, 054301 (2016).

200. M. S. Hunter, P. Fromme, Toward structure determination using membrane-protein nanocrystals and microcrystals. *Methods*. **55**, 387–404 (2011).

201. R. G. Sierra *et al.*, Concentric-flow electrokinetic injector enables serial crystallography

of ribosome and photosystem II. *Nat. Methods*. **13**, 59–62 (2016).

202. R. G. Sierra *et al.*, Nanoflow electrospinning serial femtosecond crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1584–1587 (2012).

203. D. Wang, U. Weierstall, L. Pollack, J. Spence, Double-focusing mixing jet for XFEL study of chemical kinetics. *Journal of synchrotron radiation*. **21**, 1364–1366 (2014).

204. U. Weierstall *et al.*, Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat Comms*. **5**, 3309 (2014).

205. A. Y. Lyubimov *et al.*, Capture and X-ray diffraction studies of protein microcrystals in a microfluidic trap array. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 928–940 (2015).

206. U. Weierstall, Liquid sample delivery techniques for serial femtosecond crystallography. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. **369**, 20130337–20130337 (2014).

207. W. Liu *et al.*, Serial femtosecond crystallography of G protein-coupled receptors. *Science*. **342**, 1521–1524 (2013).

208. U. Weierstall, J. C. H. Spence, R. B. Doak, Injector for scattering measurements on fully solvated biospecies. *Rev Sci Instrum*. **83**, 035108 (2012).

209. A. Aquila *et al.*, Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Opt Express*. **20**, 2706–2716 (2012).

210. J. Kern *et al.*, Methods development for diffraction and spectroscopy studies of metalloenzymes at X-ray free-electron lasers. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. **369**, 20130590–20130590 (2014).

211. K. Pande *et al.*, Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science*. **352**, 725–729 (2016).

212. C. G. Roessler *et al.*, Acoustic Injectors for Drop-On-Demand Serial Femtosecond Crystallography. *Structure*. **24**, 631–640 (2016).

213. L. C. Johansson *et al.*, Lipidic phase membrane protein serial femtosecond crystallography. *Nat. Methods*. **9**, 263–265 (2012).

214. C. Mueller *et al.*, Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Struct Dyn*. **2**, 054302 (2015).

215. M. Sugahara *et al.*, Grease matrix as a versatile carrier of proteins for serial crystallography. *Nat. Methods*. **12**, 61–63 (2015).

216. I. Johnson *et al.*, Capturing dynamics with Eiger, a fast-framing X-ray detector. *Journal of synchrotron radiation*. **19**, 1001–1005 (2012).

217. C. Broennimann *et al.*, The PILATUS 1M detector. *Journal of synchrotron radiation*. **13**, 120–130 (2006).

218. S. Botha *et al.*, Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 387–397 (2015).

219. F. Stellato *et al.*, Room-temperature macromolecular serial crystallography using synchrotron radiation. *IUCrJ*. **1**, 204–212 (2014).

220. A. E. Cohen *et al.*, Goniometer-based femtosecond crystallography with X-ray free electron lasers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17122–17127 (2014).

221. K. Hirata *et al.*, Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL. *Nat. Methods*. **11**, 734–736 (2014).

222. E. L. Baxter *et al.*, High-density grids for efficient data collection from multiple crystals. *Acta Crystallogr Sect D Struct Biol*. **72**, 2–11 (2016).

223. M. S. Hunter *et al.*, Fixed-target protein serial microcrystallography with an x-ray free electron laser. *Sci Rep*. **4**, 6026 (2014).

224. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*. **9**, 671–675 (2012).

225. T. M. McPhillips *et al.*, Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of synchrotron radiation*. **9**, 401–406 (2002).

226. N. K. Sauter *et al.*, Improved crystal orientation and physical properties from single-shot XFEL stills. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 3299–3309 (2014).

227. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

228. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

229. E. G. Kovaleva, M. S. Rogers, J. D. Lipscomb, Structural Basis for Substrate and Oxygen Activation in Homoprotocatechuate 2,3-Dioxygenase: Roles of Conserved Active Site Histidine 200. *Biochemistry-Us*. **54**, 5329–5339 (2015).

230. D. P. Deponte *et al.*, SEM imaging of liquid jets. *Micron*. **40**, 507–509 (2009).

231. M. Schmidt, K. Pande, S. Basu, J. Tenboer, Room temperature structures beyond 1.5 A by serial femtosecond crystallography. *Struct Dyn*. **2**, 041708 (2015).

232. I. Schlichting, J. Miao, Emerging opportunities in structural biology with X-ray free-

electron lasers. *Curr. Opin. Struct. Biol.* **22**, 613–626 (2012).

233. Q. Zhou *et al.*, Architecture of the synaptotagmin-SNARE machinery for neuronal exocytosis. *Nature*. **525**, 62–67 (2015).

234. A. S. Brewster *et al.*, Indexing amyloid peptide diffraction from serial femtosecond crystallography: new algorithms for sparse patterns. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 357–366 (2015).

235. D. Singh *et al.*, Structure of Escherichia coli dGTP triphosphohydrolase: a hexameric enzyme with DNA effector molecules. *J. Biol. Chem.* **290**, 10418–10429 (2015).

236. A. F. Yakunin *et al.*, The HD domain of the Escherichia coli tRNA nucleotidyltransferase has 2',3‘-cyclic phosphodiesterase, 2’-nucleotidase, and phosphatase activities. *Journal of Biological Chemistry*. **279**, 36819–36827 (2004).

237. O. B. Zeldin, S. Brockhauser, J. Bremridge, J. M. Holton, E. F. Garman, Predicting the X-ray lifetime of protein crystals. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20551–20556 (2013).

238. H. P. Stevenson *et al.*, Transmission electron microscopy for the evaluation and optimization of crystal growth. *Acta Crystallogr Sect D Struct Biol*. **72**, 603–615 (2016).

239. K. Kinkelin *et al.*, Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15277–15282 (2013).

240. F. W. Martinez-Rucobo, P. Cramer, Structural basis of transcription elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. **1829**, 9–19 (2013).

241. C. O. Barnes *et al.*, Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble. *Molecular Cell*. **59**, 258–269 (2015).

242. S. A. Seibold, B. N. Singh, C. Zhang, M. Kireeva, Conformational coupling, bridge helix dynamics and active site dehydration in catalysis by RNA polymerase. *et Biophysica Acta (BBA* (2010).

243. R. Weinzierl, The nucleotide addition cycle of RNA polymerase is controlled by two molecular hinges in the Bridge Helix domain. *BMC biology* (2010).

244. S. Sainsbury, C. Bernecky, P. Cramer, Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*. **16**, 129–143 (2015).

245. A. Y. Lyubimov *et al.*, IOTA: integration optimization, triage and analysis tool for the processing of XFEL diffraction images. *Journal of applied crystallography*. **49**, 1057–1064 (2016).

246. L. C. Andrews, H. J. Bernstein, NearTree, a data structure and a software toolkit for the nearest-neighbor problem. *Journal of applied crystallography*. **49**, 756–761 (2016).

247. O. B. Zeldin *et al.*, Data Exploration Toolkit for serial diffraction experiments. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 352–356 (2015).

248. M. Uervirojnangkoorn *et al.*, Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *eLife*. **4**, 213 (2015).

249. C. O. Barnes *et al.*, Assessment of microcrystal quality by transmission electron microscopy for efficient serial femtosecond crystallography. *Arch. Biochem. Biophys.* **602**, 61–68 (2016).

250. J.-P. Colletier *et al.*, De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature*, 1–27 (2016).

251. G. F. Schröder, M. Levitt, A. T. Brunger, Deformable elastic network refinement for low-resolution macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2241–2255 (2014).

252. B. Wang, A. V. Predeus, Z. F. Burton, M. Feig, Energetic and Structural Details of the Trigger-Loop Closing Transition in RNA Polymerase II. *Biophys. J.* **105**, 767–775 (2013).

253. D. Z. Hazelbaker, S. Marquardt, W. Wlotzka, S. Buratowski, Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Molecular Cell*. **49**, 55–66 (2013).

254. M. E. Schmitt, T. A. Brown, B. L. Trumpower, A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Research*. **18**, 3091–3092 (1990).

255. J. A. Ranish, S. Hahn, The yeast general transcription factor TFIIA is composed of two polypeptide subunits. *Journal of Biological Chemistry*. **266**, 19320–19327 (1991).

256. M. M. Mbughuni, K. K. Meier, E. Münck, J. D. Lipscomb, Substrate-mediated oxygen activation by homoprotocatechuate 2,3-dioxygenase: intermediates formed by a tyrosine 257 variant. *Biochemistry-Us.* **51**, 8743–8754 (2012).

257. S. L. Groce, J. D. Lipscomb, Aromatic ring cleavage by homoprotocatechuate 2,3-dioxygenase: role of His200 in the kinetics of interconversion of reaction cycle intermediates. *Biochemistry-Us*. **44**, 7175–7188 (2005).

258. E. G. Kovaleva, J. D. Lipscomb, Crystal structures of Fe2+ dioxygenase superoxo, alkylperoxo, and bound product intermediates. *Science*. **316**, 453–457 (2007).